Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *University of Michigan*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Effects of Discrimination Training on Stimulus Generalization: H. M. HANSON 321
Generalization Gradients around Stimuli Associated with Different Reinforcement Schedules: N. GUTTMAN 335
Force Emission during Bar Pressing: J. M. NOTTERMAN 341
Stimulus Generalization and Discrimination Learning by Primates: J. M. WARREN AND K. H. BROOKSHIRE 348
Influence of Work Distribution upon Complex Learning by the Noncorrection and Modified-Correction Methods: C. E. NOBLE AND A. TAYLOR 352
Immediate Vs. Delayed Shock in T-Maze Performance: R. K. KNAPP, R. H. KAUSE, AND C. C. PERKINS, JR. 357
CS Termination and the Response Strength Acquired by Elements of a Stimulus Complex: D. D. WICKENS AND H. A. CROSS 363
The Role of Punishment in Figure-Ground Reorganization: G. L. MANGAN 369
A Methodological Study of the Form and Latency of Eyelid Responses in Conditioning: K. W. SPENCE AND L. E. ROSS 376
The Effects of Retinal Locus and Attention on the Perception of Words: H. S. TERRACE 382
Factors Which Influence Subjective Probability: S. S. KOMORITA 386
Acquisition of Probabilistic Paired Associates as a Function of $S\text{-}R_1$, $S\text{-}R_2$ Probability: J. F. VOSS, C. P. THOMPSON, AND J. H. KEEGAN 390
Secondary and Generalized Reinforcement in Human Learning: F. H. KANFER AND J. D. MATARAZZO 400
Scales of Apparent Force: J. C. STEVENS AND J. D. MACK 405
Supplementary Report: Maintenance of a Previously Learned Concept as a Function of Partial Reinforcement: A. H. BUSS 414
Replication Report: Work and Rest as Variables in Cyclical Motor Performance: A. M. BARCH 415

American Psychological Association

Vol. 58, No. 5 November 1959

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1959 the publication cost is $30 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1959 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 58, NO. 5 NOVEMBER, 1959

EFFECTS OF DISCRIMINATION TRAINING ON STIMULUS GENERALIZATION [1]

HARLEY M. HANSON [2]

Duke University

The present study utilizes operant conditioning techniques to investigate the effects of discrimination training on the generalization gradient. Previous experiments have shown orderly wavelength generalization gradients following variable interval (VI) reinforcement of the key-pecking response in the pigeon to a monochromatic stimulus (Guttman & Kalish, 1956). The present experiment investigates whether a comparable gradient can be obtained after differential reinforcement of two monochromatic stimuli and, secondly, what changes in the gradient will take place after such discrimination training. An additional aim of the experiment is to investigate the formation of discriminations with various S^+, S^- differences.

Certain predictions concerning the postdiscrimination gradient have been forwarded by various theorists. Spence (1936; 1937; 1942), in his analysis of transposition, considers the basic mechanism in discrimination to be the interaction, i.e., algebraic summation, of excitatory and inhibitory gradients. The most important implications of a Spence-type theory for the present situation [3] are the following: (*a*) The postdiscrimination gradient will be steeper than the generalization gradient in the region of S^-. (*b*) If a complete discrimination is developed, the value of the postdiscrimination gradient will be zero at S^-. (*c*) The mode of the postdiscrimination gradient will be displaced away from S^- in relation to the mode of the generalization gradient. (*d*) The magnitude of this displacement will increase as the S^+, S^- difference is reduced. (*e*) The maximum heights of the postdiscrimination gradients will be reduced as the S^+, S^- difference is reduced. (*f*) The post-

[1] This paper is based on a dissertation submitted to the Department of Psychology, Duke University, in partial fulfillment of the requirements for the Ph.D. degree. A preliminary report appeared in Hanson (1957). This research was supported by Grant MH-1002 from the National Institute of Mental Health to Norman Guttman, who served as research adviser. Thanks are due Deborah Nickerson who prepared many of the figures and helped with calculations.

[2] Now at Merck Institute for Therapeutic Research, West Point, Pa.

[3] For the present analysis, the generalization gradient is equated to Spence's gradient of excitation, and the postdiscrimination gradient is equated to Spence's residual or summated gradient.

discrimination gradient will be a fractional part of the generalization gradient and will nowhere exceed it.

Hull (1936; 1943; 1950; 1952) formally elaborated a theory of this type, substantially following Spence's formulation, but Hull's assumption of exponential gradients leads to the important difference that no shift in the mode of the postdiscrimination gradient is predicted. Since both Spence and Hull based their theorizing on data generated in the selective response situation rather than the discriminated free-operant situation, it must be remembered that the applicability of such theories to the present experiment is qualified and approximate.

Two previous experiments, by Antoinetti (in Hull, 1952) and by Passey and Herman (1955), provide only limited information about the present problem. Both attempted to determine the postdiscrimination gradient for the brightness continuum, using rats in a simple runway. Antoinetti obtained a unidirectional generalization curve after training on a black-white discrimination. Although he reported no control curve with which the postdiscrimination gradient might be compared, from the data presented the generalization decrement from black to white appeared to be nearly zero. Passey and Herman (1955), using similar apparatus and procedure, obtained postdiscrimination gradients after training on brightness ratios of 1:16 and 1:32, the negative stimuli being the same for both groups. No statistically significant differences were found.

There are few studies which deal systematically with the relationship between the rate of formation of a discrimination and the physical interval separating stimuli which differ only in one dimension. Frick (1948) showed that the response to the negative stimulus in an operant conditioning experiment extinguished more rapidly with increasing S^+, S^- differences. In a similar experiment, using a Graham-Gagné runway, Raben (1949) found more rapid extinction of the unreinforced response with greater brightness differences.

Method

Subjects.—The *S*s were 33 experimentally naive pigeons, maintained by restricted feeding at 75% to 80% normal body weight.

Apparatus.—An automatic Skinner-type key-pecking apparatus, similar to that described by Guttman and Kalish (1956) was used. The *S*'s chamber was 12½ × 14 × 13 in. high, and was located in a dark, air-conditioned, sound-protected room. The entire experimental box was painted flat black, inside and out, with a floor of sand-blasted plexiglas. A shielded 25-w. bulb was located below the floor. The source of illumination for the key was a Bausch and Lomb grating diffraction monochromator (Model 33-86-40, with 1200 grooves/mm. grating), equipped with a 108-w. ribbon filament tungsten lamp. The exit and entrance slits were set at 5 mm., providing a band width of 16.5 mμ.

Procedure.—All *S*s were adapted, magazine trained, and conditioned to key peck according to a set schedule covering 5 days. The schedule was rigidly adhered to for all *S*s, insuring comparable treatment. The floor light was turned off after Day 3; the only source of light remaining in the box was that falling on the key (the stimulus) and the light in the reinforcement magazine, which was on only during the reinforcement cycle.

All *S*s were given VI reinforcement for 5 days following conditioning. Thirty 60-sec. stimulus-on periods, alternated with 10-sec. stimulus-off periods, were given each day. The mean interval between reinforcements, not counting stimulus-off time, was approximately 60 sec. During stimulus-off periods, a shutter operated, removing the light from the key and leaving the experimental box dark. Reinforcements were never given during a stimulus-off period. If a criterion of 500 or more responses per session was met at the end of 5 days of VI reinforcement, *S* was assigned to one of four discrimination groups or the control group. The criterion was not met in only two cases, and these *S*s

were given additional days of VI reinforcement until the 500-response criterion was reached. The four discrimination groups were trained with the same positive stimulus, 550 mμ, and differed only with respect to the negative stimulus. These discrimination groups ($N = 6$ in each group) will be henceforth designated by the negative stimulus used: 555, 560, 570 and 590 mμ.[4] The control group ($N = 8$) was given no further training before generalization testing.

During discrimination training, reinforcement of the positive stimulus was maintained on the same VI schedule previously used. The negative stimulus was never reinforced. The positive and negative stimuli were presented successively in a prearranged order. Fifteen 1-min. intervals of S^+ and 15 of S^- were presented each day. These 30 stimulus presentations comprised three blocks of 10, and within each block there were 5 positive and 5 negative stimuli. These 10 stimuli were arranged in a random sequence, taken from a table of random numbers—runs greater than three being eliminated. Six different schedules, containing a different order of stimulus presentations for each training day, were prepared, and each schedule was used with one *S* in each of the discrimination groups. During all sessions, the stimuli were changed during the 10-sec. stimulus-off period.

Discrimination training was carried out to a criterion of no-responding to the five S^- presentations within any one 10-min. block, provided that responding to S^+ continued. Training was terminated if this criterion was not met in 25 days. This occurred in only one instance, an *S* in the 555 mμ group, whose data are not included in the analysis. The hypothesis of no differences among the groups on the basis of the number of responses made on the last day of VI training was tested by analysis of variance and was accepted ($F = .47$).

The day following completion of discrimination training, and for all *S*s in the control group following VI training, tests for generalization were given. Generalization testing was carried out under conditions of extinction. The stimulus-on, stimulus-off schedule was changed to 30-sec. of stimulus-on, alternated with 10-sec.-off. Eight different generalization tests were prepared, of which six were paired with the six discrimination schedules. Each test consisted of 130 30-sec. periods, divided into 10 blocks of 13 different stimuli. The same test stimuli, 480, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, and 620 mμ, were used for all *S*s. Within each block of 13, the stimuli occurred in a different random order. On the day following the first test, and with no intervening training, an identical second generalization test was given.

[4] The 590 group contained an extra *S* which, due to a clerical error, was not tested for generalization.

Results and Discussion

Generalization.—The mean generalization functions for all groups are shown in Fig. 1. These gradients were constructed by plotting the average number of responses to each stimulus in the first and second tests combined, against the wavelength of that stimulus. The data were combined in this way because the two tests constitute successive phases of extinction without intervening training. It is clear that the postdiscrimination gradients show a uniform and orderly relationship to wavelength, similar to that obtained after simple conditioning. The postdiscrimination gradients, however, differ from the control generalization gradient in two striking ways: (*a*) all of the postdiscrimination gradients appear much higher than the control gradient, and (*b*) the various postdiscrimination gradients appear to be displaced in the direction of shorter wavelength. The right sides of the postdiscrimination gradients are ordered without inversions in relationship to the various values of S^-; this orderly relationship is preserved on the left sides of the gradients, but with smaller vertical separations between the curves.

The essential pattern presented by the postdiscrimination gradients would appear to be one in which a basic rate-wavelength function, relatively constant in form and height occupies various positions on the wavelength axis. Further analysis of the data will be guided by this hypothesis,

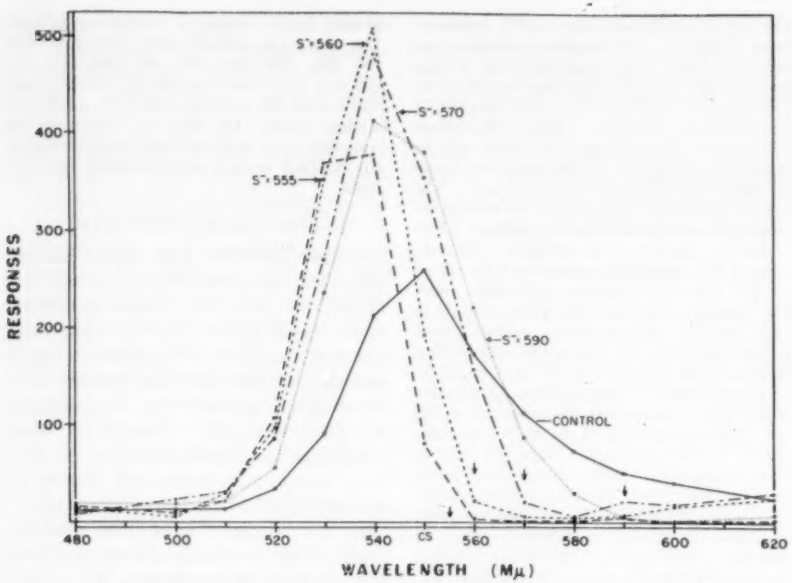


FIG. 1. Mean generalization gradients for the control and four discrimination groups, identified by the values of the negative stimulus. Arrows indicate the positions of the negative stimuli.

which is intended to be a description of the underlying curves from which the present values are sampled, not a theory as to how such functions may be generated. The departures of the present data from these hypothetical curves are regarded as the result of the particular discrete test stimulus values selected. The conspicuous truncation of the postdiscrimination gradients for the 555 and 590 mμ groups and, to a lesser extent, that of the other two groups, is presumably due to the failure to include test values close to the maxima of the underlying curves. The basic assumptions of this analysis, therefore, are capable to empirical test by the appropriate selection of test values.

The displacement of the postdiscrimination gradients was tested by determining for each *S* the percentage of the total number of responses emitted to 540 mμ and shorter wavelengths. These values are shown in Fig. 2. When tested by simple analysis of variance, the differences among the discrimination groups are significant beyond the .001 level ($F = 19.6$; $F_{.001}(3, 19\ df) = 8.28$).

The differences in height between the discrimination and the control curves, contrary to first impression, do not signify a correlated difference in total area under the curves. The total number of responses emitted in both generalization tests was taken as an estimate of the area under the function, and the between-group differences were tested by means of analysis of variance. The hypothesis of no difference in total number of responses cannot be rejected ($F = 1.25$; $F_{.05}(4, 6\ df) = 2.74$). This is consistent with the apparent similarities in form among the postdiscrimination

gradients and might be considered to suggest a redistribution of a relatively constant number of responses along the wavelength dimension as a principal effect of discrimination training.

The degree to which the group functions are representative of the functions for individual *Ss* may be demonstrated by comparing the curves in Fig. 1 with those in Fig. 3. All discrimination *Ss* without exception show what we have termed horizontal displacement of the gradient, and the magnitude of this displacement is uniform within a group. The truncation of the mean postdiscrimination gradient for the 555 mμ group (Fig. 1) is due to the fact that all *Ss* in this group show a similar truncation, and it is not the result of averaging curves centered at 530 mμ with others centered at 540 mμ. It would seem that the inclusion of 535 mμ as a test stimulus would have produced a high proportion of curves with well-defined modes for the 555 mμ group, and that the inclusion of a 545 mμ stimulus would have had a comparable effect for the 590 mμ group.

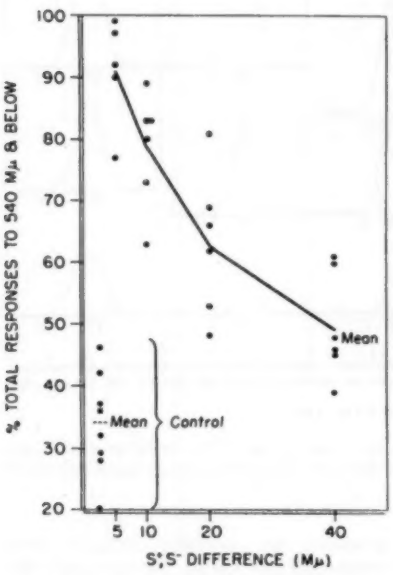


FIG. 2. Percentage of responses to wavelengths shorter than the positive stimulus, as a function of stimulus difference.

It would appear, on the basis of the foregoing results, that the following predictions derived from gradient interaction theory are supported: (*a*) increased steepness of the gradients in the region of S^-; (*b*) low response level at S^-; (*c*) displacement of the mode of the postdiscrimination gradient; and (*d*) increasing displacement of the mode with decreasing S^+, S^- difference.

However, there is a serious question as to the applicability to the present data of Predictions *e* and *f*, which mainly refer to the subtractive mechanism supposed to underly discrimination. The postdiscrimination gradients are clearly not included by the obtained control gradient. This outcome was not anticipated in the design of the experiment; the decision to terminate the training of the control group at 5 days was made because the discrimination *Ss*, even with the additional reinforcements they received, were not expected to exceed the level of the controls. However, if a control gradient of the same approximate peak height as the postdiscrimination gradients is estimated from previous data (Guttman & Kalish 1956), the postdiscrimination gradients still fail to be included.[5]

It is interesting to speculate as to what combination of positive and negative gradients would have to be assumed in order to produce the present postdiscrimination gradients by algebraic subtraction. A simplified reconstruc-

[5] The present control curve when transformed into a unidirectional gradient appears to be a member of multiplicatively related family of such curves reported by Kalish and Guttman (1957, Fig. 4). However since the present control curve was found to be significantly asymmetrical around 550 mμ by analysis of variance (violating an assumption which is basic to such a transformation), the estimated curve referred to above must be considered merely an approximation.

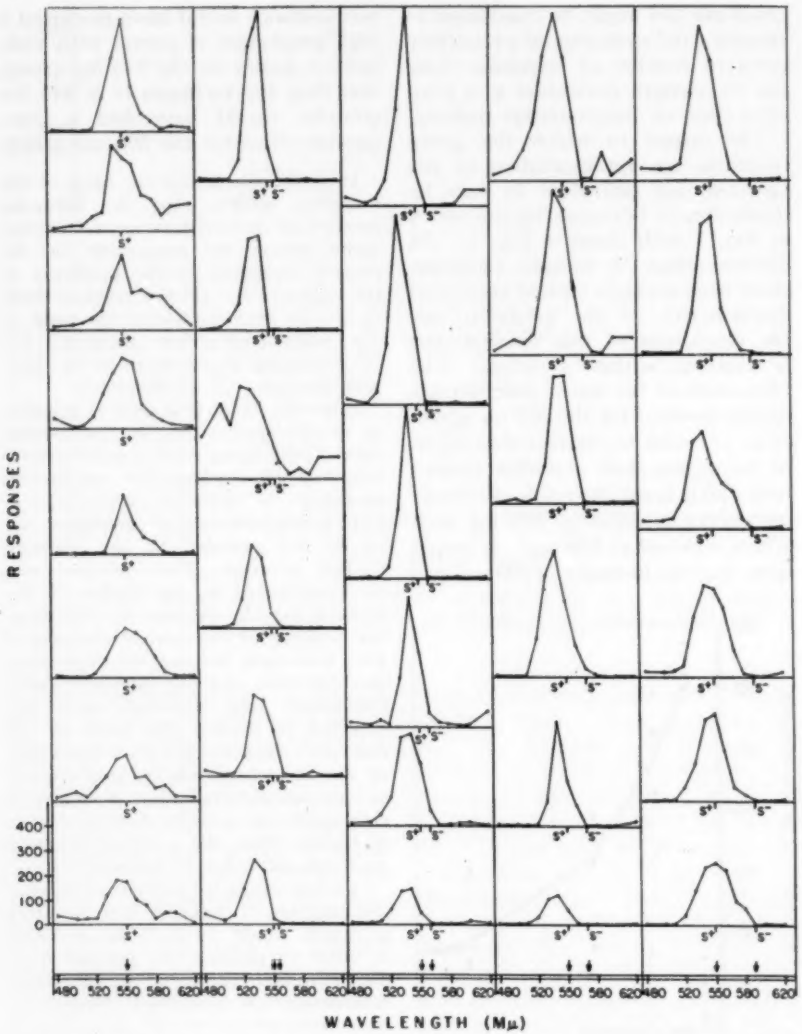


FIG. 3. Gradients obtained from individual *Ss*. The values of the training stimuli are indicated by the arrows. (The third curve from the top in the 555 mμ group represents the *S* who did not meet the criterion.)

tion (heir to all the difficulties of such simplifications) of the requisite curve forms for the 555 and 590 mμ groups is shown in Fig. 4. The curves were produced by drawing straight lines connecting the values for 520 and 530, 540 and 550 mμ for the 555 mμ group and the values for 530 and 540, 550 and

560 mμ for the 590 mμ group. The lines were extended until they intersected with each other at the "peak" of the curve and until they crossed the base line. By comparison with Fig. 2, it can be seen that these linear reconstructions essentially represent the area and form of the empirical postdiscrimination curves. Extremely conservative "excitation" curves were produced by the extension of the line representing the short wavelength side of the curves to its intersection with the abscissa value of 550 mμ. The other side of the excitation curve is merely a mirror image around 550 mμ. Hypothetical "inhibition" curves were obtained by graphical subtraction of the simplified postdiscrimination functions from the "excitation" curves. Considering such curves, the

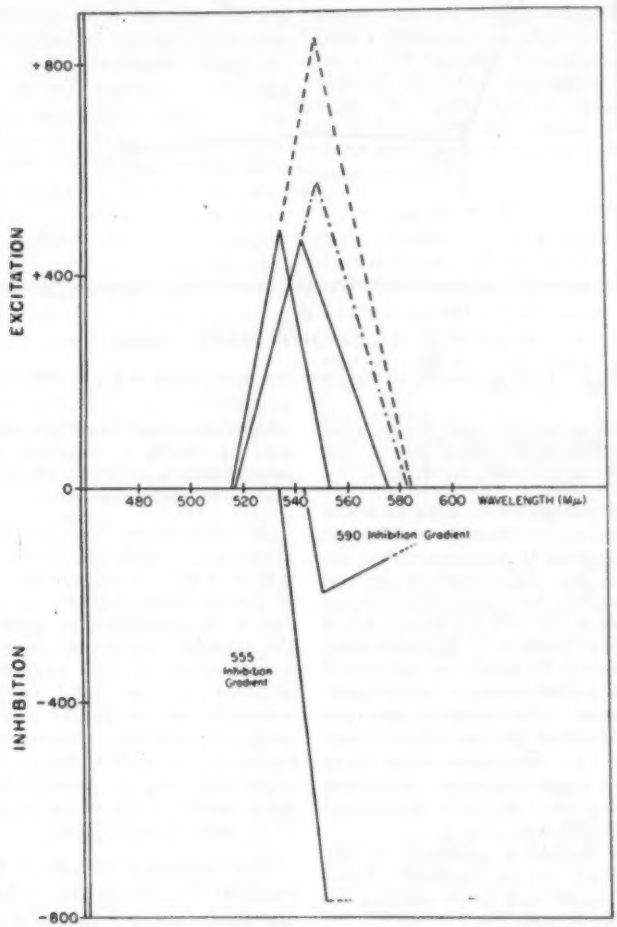


FIG. 4. Idealized postdiscrimination gradients, with the necessary theoretical excitation and inhibition curves for their derivation by subtraction.

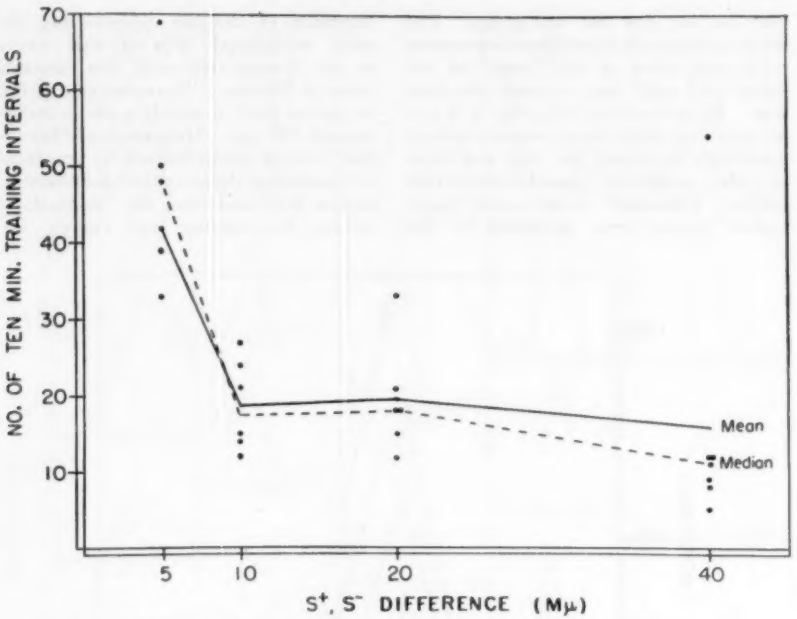


FIG. 5. Time to criterion of discrimination as a function of S^+, S^- difference.

present data apparently can be accounted for by excitation-inhibition theory, but only under the following conditions:

1. That excitations as large as would be required would be generated by the additional positive reinforcements received by the discrimination groups. This assumption seems especially tenuous in the case of the 590 mμ group, which reached the criterion of discrimination after a mean of 81 additional minutes of intermittent reinforcement on the positive stimulus. This would imply that the mean total of 231 min. of reinforcement of S^+ for this group would have generated a positive gradient more than double in area that shown by the control group after 150 min. (Fig. 1).

2. That inhibition gradients of the form depicted can be assumed. These gradients would not show maxima at the negative stimuli.

3. That all pairs of excitation and inhibition gradients would have their magnitudes so adjusted that the residual area is nearly a constant, which is approximately equal to the area under the prediscrimination curve.

An alternative, and relatively nontheoretical formulation of the foregoing results, would be to describe the effect of discrimination training as the relocation of the generalization gradient upon the stimulus continuum, together with a restriction of the stimulus region occupied by the total of responses ordinarily defined by the primary generalization gradient. No specific mechanism to accomplish these changes is postulated, and no quantitative model can presently be offered for the prediction of the form of such changes.

Discrimination training.—The time required to reach the criterion of discrimination as a function of S^+, S^- difference is plotted in Fig. 5. The differences among group means as

tested by analysis of variance are highly significant ($F = 8.13$; $F_{.001}$ (3, 20 *df*) = 8.1). The averaged curve shows a sharp increase as the negative stimulus approaches the positive in physical magnitude. The function shown is similar to those reported by Frick (1948) and Raben (1949) for similar measures. No attempt was made to fit a function to the present data because of the limited number of points available. It seems likely, however, that the intersect of the function with the wavelength axis would occur in the region of 610 or 620 mμ. With reference to Fig. 2, it might be suggested that discrimination training between 550 and 620 mμ would have a minimal displacing effect upon the postdiscrimination gradient. Similarly, it is suggested that discrimination time would be zero for stimulus values at the limit of the generalization function.

It is obvious that the function representing the relationship between training time and stimulus difference cannot intersect the ordinate at $x = 0$, but the question can be raised as to whether this function is asymptotic to zero or to some small positive value. A logical choice of a positive value would be the difference threshold in the region of the positive stimulus. Hamilton and Coleman (1933) report a difference threshold of 17 mμ at 550 mμ for the pigeon. This value is clearly too large, although it is consistent with other values of the Δλ versus λ function obtained by their method.

Since, on the basis of the present discrimination data, the difference threshold at 550 mμ is obviously less than 5 mμ, it was decided to force discrimination to its limits by successively decreasing the stimulus separation in further discrimination training with a few *Ss* after the main experiment. Although the procedure was not entirely uniform for the three *Ss* trained, the criteria of discrimination were met by all *Ss* in all problems attempted, including a stimulus difference of 1 mμ (550 mμ = S^+, 551 mμ = S^-). Training was discontinued at this point because of the limitations of the apparatus. While a terminal value of Δλ was not yielded by these data, there is little doubt that the spectral sensitivity of the pigeon for these wavelengths is in the same range as that of man (Judd, 1932). Furthermore, since the human data based on simultaneous comparison indicate a difference threshold of 1.5 to 2.0 mμ at 550 mμ, and the present training method employed successive presentations, it would not be unreasonable to suggest that the pigeon is perhaps more sensitive to wavelength differences than is man.

The course of discrimination training.—Figures 6 and 7 show cumulative response curves to the negative and positive stimuli for individual *Ss*. The negative curves show uniform negative accelerations, and the positive curves for the most part are linear. The constancy of rate of responding to the positive stimulus suggests there was little or no negative induction from the negative stimulus. In support of this notion, during the first two days of discrimination training there is a mean increase of 39% in rate of responding to 550 mμ over the rate during Day 5 of VI training. Previous experimentation[6] has shown that the rate of responding on this particular VI schedule generally increases day by day during VI training, and that the rate of increase from Day 5 to Day 6 is approximately 10%. A paired comparison *t* test for mean differences greater than 10% was computed. The *t* was 3.23 ($P < .005$). Since the mean differences exceeded the adjusted mean value, it may be concluded that the response rate to the positive stimulus increased sig-

[6] Unpublished data from N. Guttman and H. Kalish.

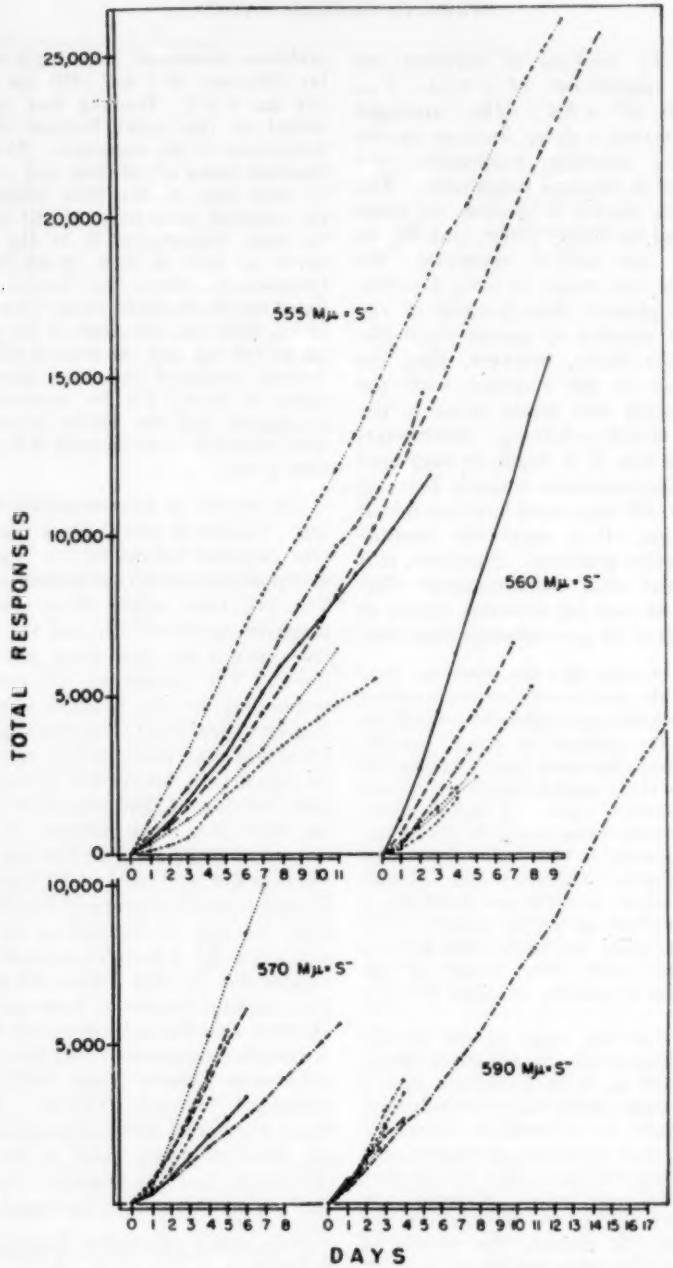


FIG. 6. Cumulative number of responses to S^+ during discrimination training.

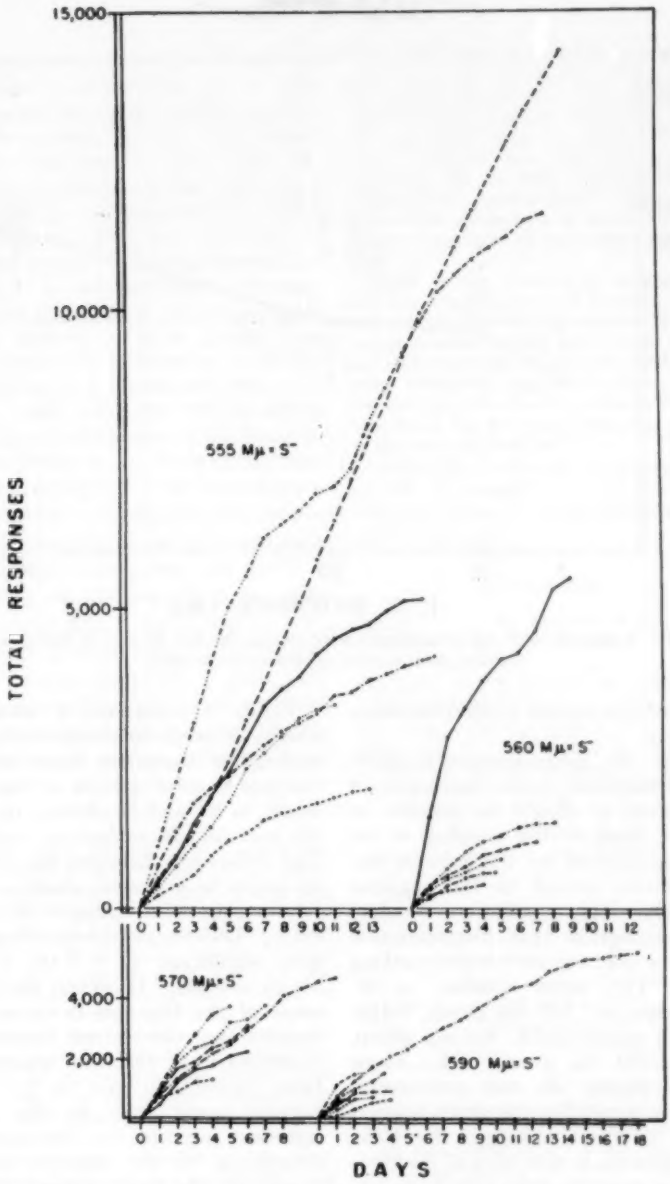


FIG. 7. Cumulative number of responses to S^- during discrimination training. The curve in the 555 mμ group which reaches 14,000 responses is that of the *S* who did not reach the discrimination criterion. Given *S*s in this and in the preceding figure are identified by the same type of line.

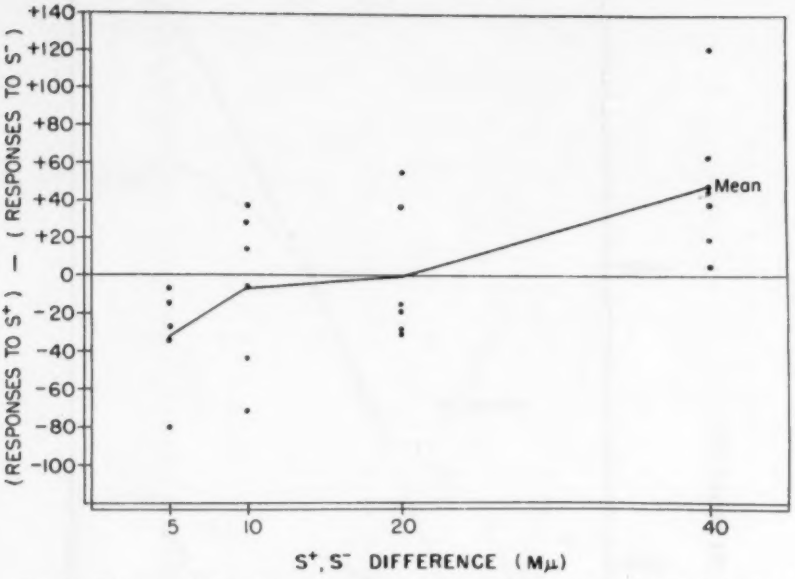


FIG. 8. Responses to S^+ minus responses to S^- during the first 10 min. of discrimination training, as a function of stimulus separation.

nificantly as a result of discrimination training.

Since the generalization gradient was determined under conditions of extinction, it should be possible to predict from it the number of responses emitted by the various discrimination groups to the negative stimulus. This prediction involves the assumption that discrimination involves only extinction and nothing more. The mean number of S^- responses are: 555 mμ group, 4,410; 560 mμ group, 1,675; 570 mμ group, 1,546; 590 mμ group, 827. These values clearly do not generate a function paralleling the generalization curve. This would suggest that some other process is operating in discrimination—possibly induction from the positive stimulus—which has a relatively larger effect for values closer to the positive stimulus.

Figure 8 represents a somewhat similar attempt to estimate the generalization decrement from the differences in total number of responses made to S^+ and S^- during the first 10 min. of discrimination training. The differences between the groups, as might be expected, shows a clear relationship to increasing S^+, S^- differences. Differences between the groups were significant ($F = 5.16$; $F_{.01}$ (3, 19 df) = 5.01). However, the placement of the function deviates from expectation, the actual number of responses to S^- for small separations being larger than that for S^+. This unusual result may be due to a spurious increase in the rate of responding to the negative stimuli caused by the time required for the consumption of the reward. In the 5 min. of responding to the positive stimulus, approximately five 3-sec.

reinforcements were given. If the lost time is accounted for in this manner for each of the groups, approximately one-third of the displacement of the function is accounted for. Another possibility is that this effect is similar to the phenomenon noted by Pavlov (1927, pp. 118–121) in the first stages of classical discrimination, i.e., an immediate decrease in reflex strength of the S^- followed by an increase up to or greater than that of the CS, followed by a regular diminution to a permanent zero.

The data from the 590 mμ group offer graphic evidence of the immediate and continuous effects of discrimination training. For the easiest problem, positive and negative discrimination curves (Fig. 6 and 7) begin with a difference in response rate which is amplified as training proceeds. The function in Fig. 8 would suggest that a similar process takes place in the other groups, although it may have been obscured by the second unidentified effect which temporarily elevates responding to the negative stimulus.

Summary

Thirty-three pigeons, divided among five groups, were trained in a Skinner-type situation. After 5 days of key-peck training with a monochromatic light of 550 mμ on the key (approximately 2.5 hr.), four of the groups were given discrimination training between 550 mμ and a negative stimulus of 555, 560, 570, or 590 mμ, one group being assigned to each value. After the criterion of discrimination training was reached, and as soon as the control group completed its preliminary training, all groups were given generalization tests, consisting of repeated presentations of permuted orders of 13 stimuli of different wavelengths.

Highly orderly generalization gradients (number of responses vs. wavelength) were obtained for all groups and individual *Ss*. The postdiscrimination gradients were found to have statistically significant differences in location on the wavelength continuum, these differences being describable as a displacement away from the negative stimulus in the direction of the positive. Although the postdiscrimination gradients were higher than the control gradient, the areas under the curves, estimated from the total number of responses, did not differ significantly. The results of the present experiment were compared with an analysis of discrimination in terms of gradients of excitation and inhibition.

Time to the criterion of discrimination was found to decrease with increasing S^+, S^- differences. The S^+ cumulative response curves obtained during discrimination training were essentially linear with a slight upward curvature, suggesting a minimum of negative induction from S^-, although evidence was found for induction from the positive to the negative stimulus.

Exploration of the limit of discriminability at 550 mμ revealed that the wavelength difference threshold in this region was not greater than 1 mμ.

References

Frick, F. C. An analysis of an operant discrimination. *J. Psychol.*, 1948, **26**, 93–123

Guttman, N., & Kalish, H. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hamilton, W. F., & Coleman, T. B. Trichromatic vision in the pigeon as illustrated by the spectral hue discrimination curve. *J. comp. Psychol.*, 1933, **15**, 183–191.

Hanson, H. M. Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*, 1957, **125**, 888–889.

Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1936, **46**, 9–30.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

Hull, C. L. Simple qualitative discriminative learning. *Psychol. Rev.*, 1950, **57**, 303–313.

Hull, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

Judd, D. B. Chromaticity sensibility to stimulus differences. *J. opt. Soc. Amer.*, 1932, **22**, 72–108.

Kalish, H. I., & Guttman, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

PASSEY, G. E., & HERMAN, P. N. The shape of the discrimination gradient for two intracontinuum stimulus separations. *J. exp. Psychol.*, 1955, **49**, 273–277.

PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep). London: Oxford Univer. Press, 1927.

RABEN, M. W. The white rat's discrimination of differences in intensity of illumination measured by a running response. *J. comp. physiol. Psychol.*, 1949, **42**, 254–272.

SPENCE, K. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

SPENCE, K. The differential response in animals to stimuli varying in a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

SPENCE, K. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 257–271.

(Received September 3, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

GENERALIZATION GRADIENTS AROUND STIMULI ASSOCIATED WITH DIFFERENT REINFORCEMENT SCHEDULES[1]

NORMAN GUTTMAN

Duke University

In this experiment, the pigeon is reinforced on a 1-min. variable interval schedule for responding to one wavelength of light and on a 5-min. variable interval (VI) schedule for responding to a different wavelength. Then the wavelength generalization gradient is obtained. The analysis of this type of training situation and of the resulting pattern of stimulus generalization is of considerable theoretical interest, because it contains features in common with several problems which have been hitherto unrelated.

First, the training situation is of the sort which Skinner (1953) has studied under the heading of the stimulus control of reinforcement schedules. He has shown that different stimuli associated with different reinforcement contingencies can come to control the rate and temporal distribution of operant behavior. If, for example, a pigeon has been trained on a variable-interval schedule with a green key and on a fixed-interval schedule with a red key, the rate of responding will be uniform in the presence of green, but in the presence of red, the rate will undergo a systematic increase from an initial low value after reinforcement (the characteristic "fixed-interval scallop"). Skinner has not dealt with these cases in relation to stimulus generalization but he has placed emphasis on the independence of the performance patterns under various schedules. The degree and conditions of such independence can, however, be explored by means of a stimulus generalization test, as will be seen.

From the point of view of Hull (1943), a rather explicit prediction can be made concerning the pattern of stimulus generalization associated with the present training conditions. In this framework one would be inclined to describe the training as unequal reinforcement of two stimuli in the same continuum. One would expect that a gradient of excitatory tendency would be generated around each stimulus, and the total generalization gradient would be a bimodal composite of the higher gradient for the strongly reinforced stimulus and the lower gradient for the weakly reinforced stimulus. It has already been shown (Bilodeau, Brown, & Meryman, 1956; Kalish & Guttman, 1957) that bimodal generalization gradients are obtained when two stimuli are equally reinforced.

Another set of expectations concerning the form of the generalization gradient to be obtained under these conditions derives from a recent study by Hanson (1956) dealing with the generalization gradient after discrimination training. Hanson has shown that when one wavelength is reinforced on a VI schedule and another

[1] The basic idea of this experiment was contributed by Deborah H. Nickerson. The author wishes to acknowledge the aid of Renate Guttman in carrying out the experiment and analyzing the data. This study was supported by Grant M-1002 from the National Institute of Mental Health, U. S. Public Health Service, and by the Duke University Research Council.

wavelength is extinguished, the resulting postdiscrimination gradient exhibits: (*a*) virtually zero rate in the region of the negative stimulus, (*b*) a rise to an *intermediate* rate in the region of the positive stimulus, (*c*) a continued rise to a new maximum beyond the positive stimulus, and (*d*) then a decline. This pattern of generalization, notably the phenomenon of "peak shift," was obtained for all *S*s in the Hanson experiment and has since been confirmed in several replications of procedure (e.g., Honig, 1958; Honig, Thomas, & Guttman, 1959). This type of gradient appears to be the distinctive mark of a successive[2] wavelength discrimination in the pigeon.

However, it may be seen that the training procedure of the present study is intermediate between that in the Hanson study, where one wavelength was never reinforced and another reinforced intermittently, and that of the Kalish and Guttman (1957) study, where two wavelengths were reinforced on the same schedule. If reinforcement schedules of various frequencies are on a continuum between uniform reinforcement and extinction, we can then expect the present training procedure to yield something that resembles a postdiscrimination gradient. Such a result would assist in the behavioral specification of these various types of training situations and provide us with clues as to their common features.

Method

Subjects.—The *S*s were seven experimentally naive White Carneau pigeons, maintained by restricted feeding to 80% of normal body weight.

Apparatus.—Two automatic Skinner boxes similar to those previously described (Guttman & Kalish, 1956) were used. The boxes were of identical construction except for the sources of monochromatic key illumination. One box, which was used for half of the training of all *S*s and all of the generalization testing, was fitted with a Bausch and Lomb diffraction-grating monochromator. The other box, which was used for the other half of the training, employed two Bausch and Lomb interference filters on a swinging mount. One filter was of 550 mμ peak transmission. The other was of 570 mμ and was coupled with a Wratten K2 filter to mask out short wavelengths. Neutral density filters were used to match the intensities of the filter-produced colors to the monochromator colors. Energy measurements were made by means of a photoelectric transducer with a photomultiplier pickup. No light was present in the Skinner boxes except that transmitted by the key and that accompanying the opening of the food magazine. White noise was constantly present in the animal's chamber.

Procedure.—Training to eat from the food magazine was carried out on the first day of training, and conditioning of the key-pecking response was accomplished on the following day. Four *S*s were conditioned to 570 mμ and three to 550 mμ. Fifty continuous reinforcements were obtained by all *S*s on this day. Following acquisition, 12 daily sessions of variable interval reinforcement were carried out. Each session consisted of 20 min. of training, and successive minutes were separated by 10-sec. blackout periods. During each session two 5-min. blocks of 550 mμ key illumination were alternated with two 5-min. blocks of 570 mμ key illumination. When 550 mμ was present on the key, reinforcement was on a variable interval schedule with a 1-min. mean inter-reinforcement interval (VI-1). When 570 mμ was presented, reinforcement was on a 5-min. VI schedule (VI-5). On the first day of training for all *S*s, the 5-min. stimulus blocks were presented in the order: 550–570–550–570. The second day began with 570 mμ, and this alternating order of stimuli within days and between training days was continued to the end of training.

On the day following the completion of VI training, a wavelength generalization test was administered under experimental extinction. Each of the following wavelengths was presented 12 times: 510, 520, 530, 540, 550, 555, 560, 570, 580, and 590 mμ. These 10 stimuli were arranged in 12 different random orders for each *S*. Each stimulus presenta-

[2] The distinction between successive and simultaneous training is an important one, since Honig (1958) has shown that a different form of generalization gradient is found after simultaneous discrimination.

tion was of 30-sec. duration, preceded by a 10-sec. blackout.

Results and Discussion

Performance under VI schedules.—Definite and consistent differences in rate of responding were established by the training procedure. The mean rate to 550 mμ, the VI-1 stimulus, exceeded the rate to 570 mμ on all 12 days of training, as shown in Fig. 1. Of the 84 training sessions (7 *Ss* × 12 days), 71 showed 550 mμ rate greater than 570, and all *Ss* gave more total responses to 550 mμ. The rate to 550 mμ rises until the end of training, while the 570 mμ rate appears to reach stability about midway through the course of training. Overall, 67.5% of all responses were given to the more strongly reinforced stimulus. This percentage shows a tendency to increase, not because of the extinction of 570 mμ responding, but because of the absolute increase in 550 mμ responding.

Stimulus generalization.—The mean wavelength generalization gradient obtained from the present *Ss* is shown in Fig. 2. The curve is clearly a unimodal rather than a bimodal function of wavelength, and the maximum rate of responding is found at 540 mμ rather than at the wavelength of either conditioned stimulus. All *Ss* yielded curves similar to the mean gradient, and the maxima for individual *Ss* were all at 545 mμ or below.

The form of the present gradient is nearly identical to the one produced by that group in Hanson's study which learned a complete discrimination between the same pair of wavelengths. The gradient for Hanson's group is plotted in Fig. 2 for comparison. This curve has been multiplied throughout by the ratio of the two

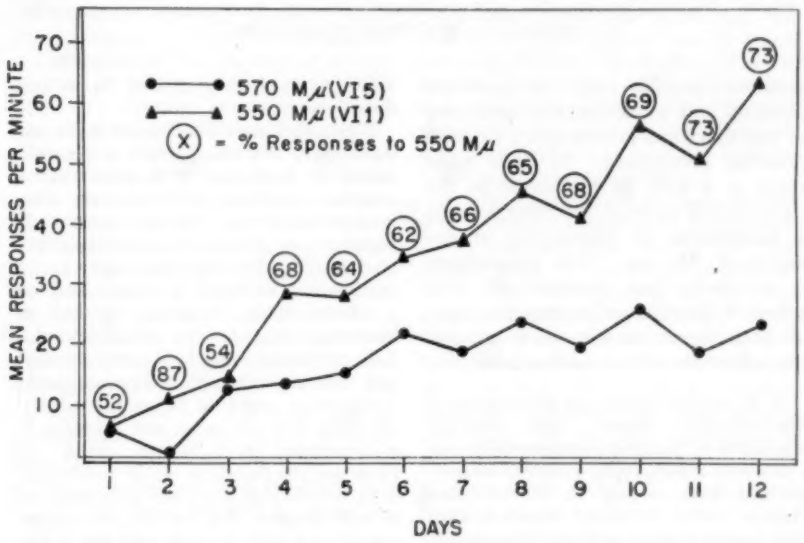


Fig. 1. Acquisition curves for key pecking in the presence of two stimuli correlated with different reinforcement schedules.

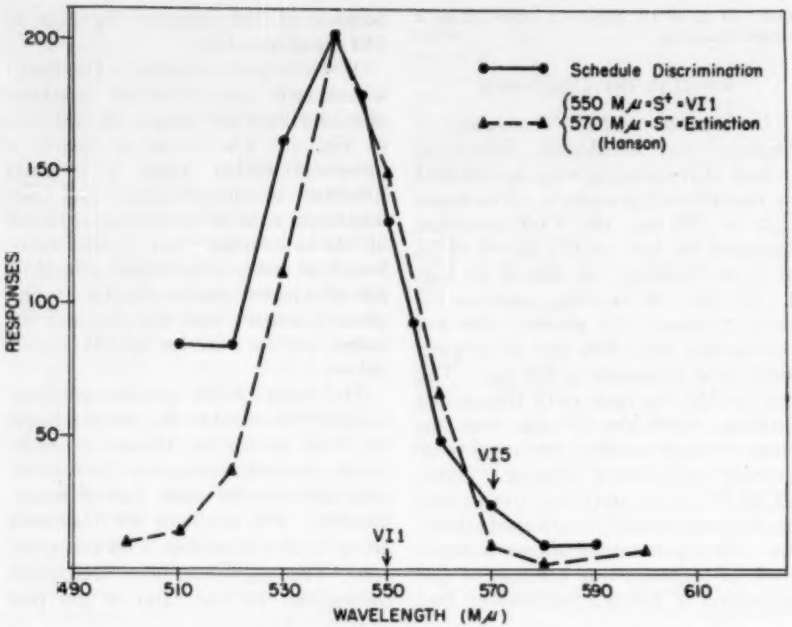


FIG. 2. Stimulus generalization gradients following schedule discrimination and successive discrimination to a criterion of extinction.

maxima (at 540 mμ) to facilitate plotting and to exhibit the similarity of the functions produced by the two training procedures. In both cases there is a shift of the mode in the direction of shorter wavelengths and a suppression of responding in the region of 570 mμ. The suppression is relatively less pronounced after schedule discrimination, but the overall patterns are so alike as to suggest the operation of one mechanism.

It is evident that the association of discriminative stimuli with different schedules of positive reinforcement can produce in the organism a state of responsiveness with respect to the stimulus domain which is nearly identical with that resulting from ordinary discrimination. The conceptual and theoretical consequences of this finding are considerable, and they extend in several directions.

First, and somewhat apart from the question of how this pattern of generalization is produced, it is clear that a stimulus associated with *relatively* weak reinforcement can become *functionally negative*. In general, it seems intuitively obvious that the algebraic sign of any conditioned stimulus is determined on a relative basis. Suppose, by way of illustration, that for one animal Stimulus A is correlated with food reinforcement and Stimulus B with no-consequences (extinction), and that for another animal Stimulus B is the same and Stimulus C is correlated with electric shock. If these are the complete set of conditions, it is indubitable that the properties of B will depend not on its zero-consequences as such, but on whether it has been alternated with A or with C. The operation of a "frame of reference"

provided by the alternate condition is unexpected in the present case because a 5-min. VI schedule is ordinarily sufficient to sustain rather strong behavior, and it evidently does so under the present training conditions. The negativity of the stimulus associated with this schedule only becomes manifest under the altered conditions of a generalization test.

It is difficult to escape the view that stimuli associated with relatively weak reinforcement and those associated with extinction both possess suppressive or inhibitory functions, as distinct from their mere correlation with observed states of weak or absent behavior. Observably, the tendency to respond to 570 mμ in the present experiment undergoes nothing but an *increase* during training, yet during generalization testing there is an absence of responding to this and adjacent stimuli. The obvious inference is that during training the behavior in the presence of this stimulus is under an active restraint, and that the observed level of performance is the resultant of this discriminative restraint and the positive factor of reinforcement. There are two sources of reinforcement for responding in the presence of 570 mμ: direct (the reinforcements received in the presence of this stimulus), and indirect (by generalization from 550 mμ). During generalization testing, both sources of reinforcement are removed and the latent inhibitory factor becomes emphasized.

If the attribution of inhibitory properties to the weaker-schedule stimulus is accepted, it may be possible to deal with the shifting of the mode of the generalization gradient by means of a model similar to, but not identical with, that suggested by Spence (1937) in his classical analysis of the transposition phenomenon. Spence's theory does indeed imply a displacement of the mode of the postdiscrimination gradient, but it does not seem feasible to apply his model to the case in point without certain modifications. In the model Spence has proposed, one of the several necessary conditions for obtaining a peak-shift is that the net behavior tendency to the negative stimulus falls "below zero." If one were to assess the state of the response to 570 mμ from the observed rate during training, one would have to draw the gradient around this stimulus as positive rather than negative. Furthermore, all of the considerations adduced by Hanson (1956) for the inapplicability of the Spence model to the results of ordinary discrimination would appear to be relevant in the present case. (The reader is referred to Hanson's treatment, which need not be reproduced here.)

The present observations do not permit us to discriminate between alternative views of the suppressive character of negative stimuli. This function may be conceptualized either as that of evoking a specific (if idiosyncratic) response which is in competition with the reinforced response under observation, or that of generating an inhibitory state which is not further specified. The important fact, however, is that either of these interpretations is distinguishable from the view that the process of discrimination can be essentially described by the weakening of the response under observation, and that a response so weakened then becomes passive in the system. The evidence here presented strongly argues that the latter description is insufficient.

A final implication of the present study is that behaviors controlled by reinforcement schedules are limited in their independence by the degree of generalization that exists among the schedule-associated exteroceptive stimuli. To facilitate a more systematic and complete approach to the problem of independence, it may be useful to consider schedule-controlled behaviors as potentially linked by stimulus similarities of various sorts: those explicitly exteroceptive (here, adjacent photic wavelengths), those explicitly proprioceptive (derivative from similar movement patterns), and those dependent upon the temporal arrangements of reinforcement

per se, and probably in some manner interoceptive.

Summary

Seven pigeons were given variable interval reinforcement for pecking at two wavelengths of light presented alternately. Reinforcement was on a 1-min. schedule in the presence of 550 mμ and on a 5-min. schedule in the presence of 570 mμ, and the rate of responding became consistently higher under 550 mμ. A test of stimulus generalization in the wavelength continuum revealed a gradient essentially the same as that obtained by Hanson (1956) after discrimination training involving complete extinction of responding to 570 mμ and intermittent reinforcement of 550 mμ.

It is concluded that the stimulus associated with the weaker of two reinforcement schedules possesses inhibitory properties in common with the extinguished stimulus of the ordinary discrimination situation. It does not appear feasible to describe the obtained gradient as the summation of a stronger and a weaker excitation, nor to account for the displacement of the gradient mode by a gradient of extinction around the "negative" stimulus, since this stimulus is negative only in a relative sense.

References

Bilodeau, E. A., Brown, J. S., & Meryman, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51**, 293–298.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hanson, H. M. The effects of discrimination training on stimulus generalization. Unpublished doctoral dissertation, Duke Univer., 1956.

Hanson, H. M. Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*, 1957, **125**, 888–889.

Honig, W. K. Prediction of preference, transposition, and transposition-reversal from the generalization gradient. Unpublished doctoral dissertation, Duke Univer., 1958.

Honig, W. K., Thomas, D. R., & Guttman, N. Differential effects of massed extinction and discrimination training on the generalization gradient. *J. exp. Psychol.*, 1959, **58**, 145–152.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Kalish, H. I., & Guttman, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

Skinner, B. F. Some contributions of an experimental analysis of behavior to psychology as a whole. *Amer. Psychologist*, 1953, **8**, 69–78.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received October 24, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

FORCE EMISSION DURING BAR PRESSING[1]

J. M. NOTTERMAN

Princeton University

This experiment was concerned with a detailed description of the emission of forces by rats in a substantially conventional Skinner box as the animals proceeded through various stages of the learning process (specifically, operant level determinations, regular reinforcement, and extinction) without any experimenter-manipulated exteroceptive discriminative stimuli.

Rudimentary data of this type were obtained by Skinner (1938, p. 312), but they are—for present purposes—deficient in two respects: the measure of force used by Skinner was not linear with actual force, and the technique employed did not permit measuring the force of single responses. A subsequent major contribution to the systematic description of free force emission was made by Trotter (1956). Unfortunately, although he apparently had the instrumentation capability, Trotter failed to keep records during the acquisition phase of a knob-pressing operant, thereby precluding the kind of response-by-response analysis presented here. Nonetheless, his comments concerning "the physical properties of bar pressing behaviour" remain the clearest exposition of the physics of bar pressing to date.

METHOD

Apparatus.[2]—The general apparatus here described provides an analysis of three intensive aspects of bar pressing behavior: force, duration, and time integral of force. The foregoing is achieved without sacrifice of the customary pulse recording techniques, simultaneous measures of digital and continuous response characteristics being possible with this system. In addition, and in anticipation of future research requirements, the same apparatus permits selective reinforcement in proportion to the magnitude of the particular response characteristic sensed.

A standard small animal cage is equipped with a modified bar, the innovation consisting in replacement of the microswitch with a pair of strain gauges. Pressure on the bar produces a minute change in the internal resistance of the strain gauges, thereby unbalancing a Wheatstone bridge to which the strain gauges are coupled. The pressure-proportional voltage thus made available is passed into a preamplifier, from which it enters an analog computer. The computer has several functions, some of which are mutually exclusive, these being exercised at *E*'s discretion. Functions pertinent to the investigation here reported are as follows: (*a*) determination of the constant number of pellets by which each suprathreshold response is to be reinforced; (*b*) further amplification of the voltage originating from pressure on the bar, thereby producing a voltage-proportional force readout on a strip-chart recorder; (*c*) generation of reference voltages used to determine the "response threshold," or the precise level of force required to activate the reinforcement circuit (accomplished jointly with a "Threshold Adjust" unit); and (*d*) determination of occurrence of suprathreshold responses on a "Yes-No" basis, for pulse counting and cumulative recording purposes.

As Functions *c* and *d* indicate, it is possible to obtain conventional operant behavior data, with the Threshold Adjust being set to some value corresponding to the force required to close the microswitch in the usual bar. It should be noted, however, that although the

[1] This research, and the program of which it is a part, is supported by the Office of Naval Research, Physiological Psychology Branch, under Contract NONR 1858 (19).

[2] The electronic design and computer programming were supervised by N. B. Marple, A. F. Sciorra, and D. Brown, Electronics Research Laboratories, Columbia University. Such electronic sophistication as may appear in the apparatus description should be attributed to these engineers, not to the writer.

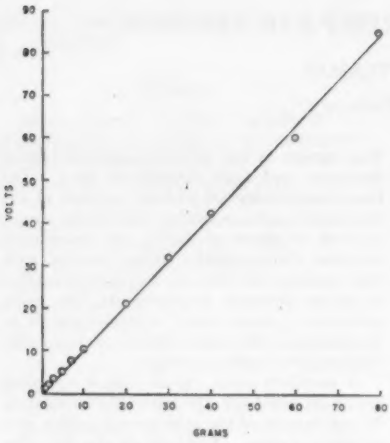


FIG. 1. Linearity check: voltage output as a function of grams weight of force on lever.

visual appearance of that portion of the bar which is within the experimental cage is identical with the conventional manipulandum, the tactual bar is quite different. In the present case, the bar remains more or less immobile during active responding (maximum movement being approximately 1–2 mm.), and was designed this way in order to lay the groundwork for potential experimental separation of movement from pressure cues (or "work" from "impulse").

Figure 1 is a linearity and calibration check between gram weights on the bar, and voltage output of the computer (Function *b*). As can be seen, the system is linear over a large range of forces. Similar checks with similar results have been obtained for time integral of force (gm.-sec.), and for the excursion of the recorder pens.

Figure 2 is a sample record. For this experiment, the critical threshold for reinforcement was set at 3 gm., this being well above the noise level of the system, but still low enough to provide *S* with essentially "free" lower and upper force limits. A response was defined as being in effect from the instant *S* emitted 3 gm. of force to the instant the force level went below 3 gm., as in "A." In the event that, during continuous contact with the bar, the force exceeded 3 gm., fell off to less than 3 gm., and then again increased past the critical threshold, two responses were counted ("B"). If the force failed to reach 3 gm., as in "C," it was not deemed a response.

The "Reinforcement Indicator" channel served the function of providing a convenient check on when the force level fell below the critical threshold. Each mark indicates delivery of one pellet, and corresponds to the instant that the force drops below critical threshold. (Duration and time integral of force data, although recorded, are not considered in the present paper.)

The foregoing convention concerning definition of a response was adopted in this experiment on the basis of its being most analogous to that of the usual bar pressing situation. (Other response criteria, e.g., calling each force emission a response, regardless of whether it reaches critical threshold are, of course, possible, and are currently being examined.) "B" is equivalent to the case in which *S* presses twice without releasing the bar, and "C" is similar to a bar contact not strong enough to close the microswitch. In general, then, the present bar is chiefly distinguished from the usual bar by its being much more "sensitive," and by its being an analog, rather than make-break, type of manipulandum.

Procedure.—Six male Wistar rats, approximately 90 days old, were placed on 22-hr. hunger rhythm for 10 days prior to operant level determinations. The experiment proper was conducted over eight successive days (or sessions), as follows: *Operant Level*, two 35-min. sessions, followed by 20 min. of tray approach training; *Conditioning*, four sessions of regular reinforcement, one pellet per response, each session terminating upon procurement of approximately 50 pellets; *Extinction*, two 35-min. sessions. Pellets were standard Noyes Co., 45 mg. each. The *S*s were fed Purina Chow for 1 hr. following each experimental session, and then again placed on deprivation.

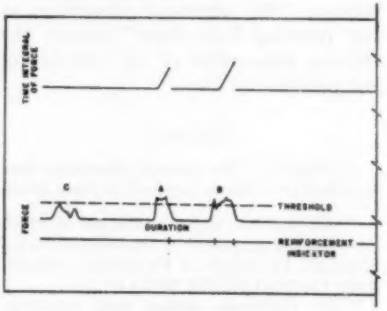


FIG. 2. Sample record.

Results and Discussion

Figure 3 is a sequential plot of peak force attained by S_1 for each of the suprathreshold bar presses emitted during the entire course of the experiment. The data shown are characteristic of the remaining five *S*s.

Noteworthy is the obvious drop in magnitude of response during the regular reinforcement sessions; this is accompanied by a decrease in response variability, or the development of "stereotypy." It is not surprising that the level of force characteristic of regular reinforcement is apparently related to the critical threshold. Skinner's data, (1938, p. 312) for example, show that his *S* came to respond with a force of some 35–40 gm.; this for a bar which apparently required almost 20 gm. of force for activation. The *S*s in this study come to press typically, during regular reinforcement, with approximately 5–6 gm.; but the critical force required was only 3 gm. Obviously, this is the result of the same sort of organism-mediated force differentiation which Skinner postulates on the basis of his data. The intriguing possibility exists that, over a significant range, *S*s will stabilize during regular reinforcement at a force magnitude which is roughly twice that of the critical threshold, as determined by *S*'s force discrimination difference limen.

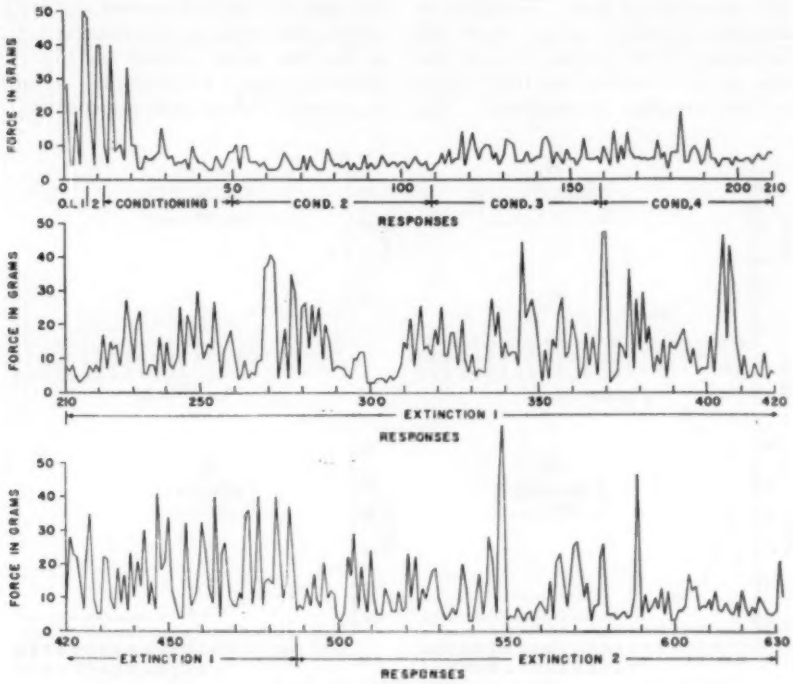


FIG. 3. Sequential emission of peak forces during successive phases of the experiment. Data are for S_1.

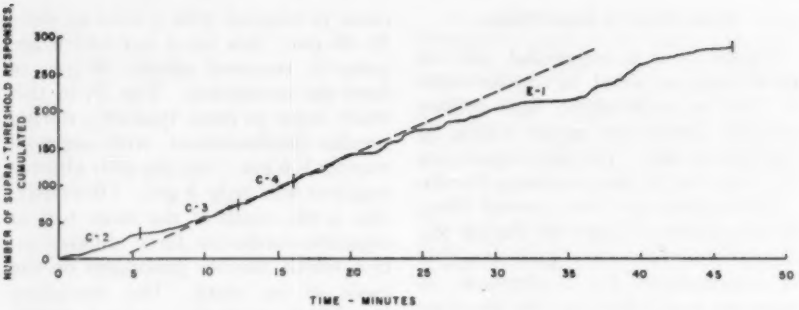


FIG. 4. Cumulative response record for S_1, same responses, as in Fig. 3.

Shortly after extinction is begun, both force magnitude and variability show a sharp increase. The trial-by-trial extinction data of Fig. 3 may be in conflict with Skinner's (1938, p. 313) observation that: "Stronger responses generally occur near the beginning of the extinction and give way to an unusually low force which is then steadily maintained." The present data, as represented by S_1, reveal a tendency for emission of fairly high magnitudes of forces well into extinction.

Figure 4 is a tracing of the usual cumulative response record obtained during bar pressing experiments, and is for the same animal as Fig. 3. (Operant Level, Conditioning 1, and Extinction 2 data were omitted from

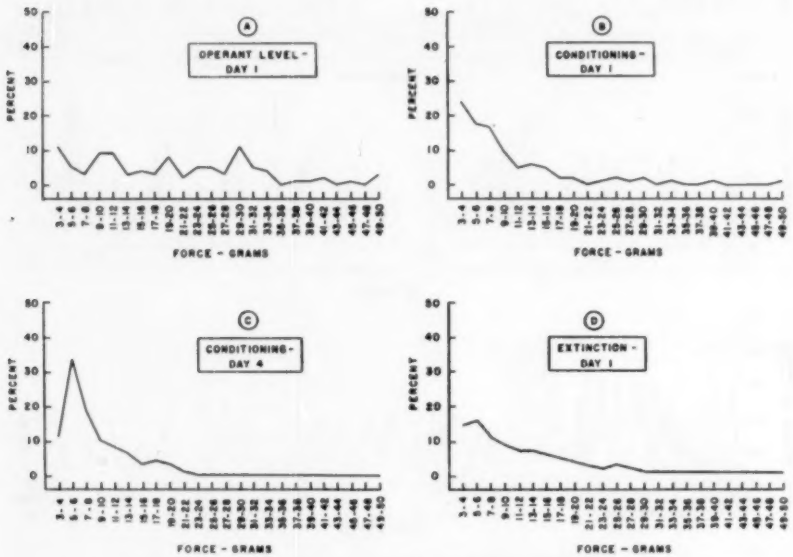


FIG. 5. Group frequency distributions of peak forces.

this figure in order to permit as large a reproduction scale as possible.) The broken line represents the slope of Conditioning 4, and indicates that this phase produced the highest over-all rate. The early stages of Extinction 1 show the customary, momentary increases in rate, followed by an over-all decline. It should be noted that the cumulative recorder's mechanism was somewhat slower than that of the strip-chart recorder; hence, the latter occasionally indicated responses which were not sensed by the former. In all computations involving number of responses, the strip-chart data were considered definitive.

The four frequency distributions comprising Fig. 5 show, in terms of percentage of total number of responses for each of the indicated experimental sessions, how the peak forces were distributed by the indicated scale intervals. These data are based on all six *S*s; individual plots are quite similar. Most striking is the shift in distribution of forces as *S*s proceed through the various phases of the experiment. The Operant Level data ("A") show quite clearly that prior to the response differentiation which takes place during conditioning, *S*s exhibit a wide range of "freely" emitted forces. Conditioning 1 ("B") depicts the beginnings of a differentiation which is fairly sharp by the end of the fourth conditioning session ("C"). Noteworthy in connection with the latter is the decrease in frequency of occurrence of 3–4 gm. forces. The *S*s apparently come to "play it safe," and typically emit forces at a level determined at the lower end by the critical threshold,

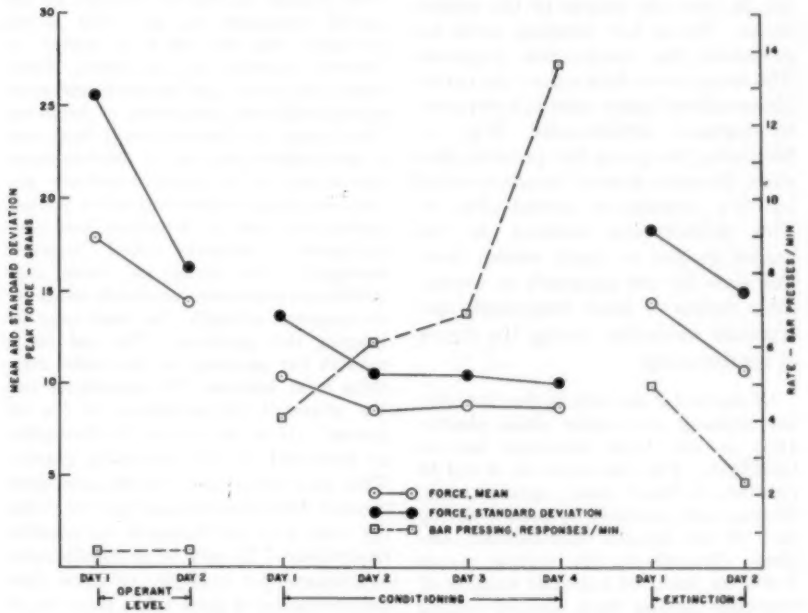


FIG. 6. Comparison of force (mean and *SD*) and rate of bar pressing.

and at the upper end perhaps by the negative aspects of unnecessary force expenditure. The latter speculation is, of course, intimately related to past notions of "least effort" or "work," and it is hoped that later analyses based upon the time integral of force measure (or "impulse") will permit careful sifting of that which may be of value in these earlier formulations.

The Extinction 1 data ("D") reveal the beginning of a return to the type of distribution characteristic of Operant Level. Whether the "random" nature of the latter is actually ever reached is presently uncertain. (The relation of force variability to Schoenfeld's [1950] and Antonitis' [1951] prior work in stereotyping has been investigated recently by Goldberg [1959].)

Figure 6 gives the group means and *SD*s of the peak forces emitted by all six *S*s over the course of the experiment. Mean bar pressing rates are presented for comparison purposes. The mean force data reflect the earlier observations based upon S_1's response-by-response performance (Fig. 3). Similarly, the group bar pressing data show the same general trends revealed by S_1's cumulative record (Fig. 4). The stereotyping evident in the earlier graphs is again shown here, this time by the approach to asymptotic values of force magnitude and standard deviation during the course of conditioning.

Of particular interest, is the fact that bar pressing rate keeps rising sharply even as the force measures become stabilized. The rate measure, it will be recalled, is based upon approximately 50 responses produced by each *S* during each of the regular reinforcement sessions. Accordingly, the increase in rate is *not* the result of a greater number of responses having been emitted during successive, constant total time sessions, but rather the result of the same number of responses having been emitted during progressively shorter sessions. Such being the case, it is clear that it is the dropping out of intervening behavior (i.e., behavior between successive bar presses) which is responsible for the observed rate increases. Considered in this sense, bar pressing rate is a measure of bar pressing response strength *relative* to the strength of other ongoing behavior. As Skinner puts it: ". . . the main datum to be measured in the study of the dynamic laws of an operant is the length of time elapsing between a response and the response immediately preceding it or, in other words, the rate of responding" (Skinner, 1938, p. 58).

The point of the preceding comment is that bar pressing rate is inherently a composite measure of a given set of behavioral events separated in time by other (albeit chain-related) behavioral events. The intensive attributes of response, on the other hand, are by their very nature exclusively indices of the actual responses per se. This is not to imply that the one is a "better" or "worse" measure than the other. Obviously, frequency and intensity measures sample different properties of behavior. The former has been shown to bear such a predictable relation to reinforcement operations as to stand relatively unchallenged as a convenient index of these operations, and by definition has come to signify a property called "response strength." The latter, i.e., force, may as Skinner says bear "no simple relation" to response strength, but that may be begging the question. The real issue, as with bar pressing, is: Do lawful relations exist between *E*'s operations and the intensive characteristics of *S*'s response? Here the answer is affirmative as indicated by the foregoing graphs. That such lawfulness is predicated upon response differentiation, and not bar pressing rate, does not diminish its possible importance. It may be of considerable significance, for example, to know that differentiation of force takes place much sooner than stabilization of bar pressing

rate (Fig. 6), and that this differentiation is maintained at stable levels, even though bar pressing occurs at more frequent intervals, indicating the gradual disappearance of other ongoing behavior.

Summary

A detailed description was made of the emission of forces by rats during acquisition (regular reinforcement) and extinction of the bar pressing response. It was found that the distribution of forces emitted during acquisition peaked at a value approximately twice that of the force required for reinforcement, and that both magnitude and variability of force decreased during acquisition and increased during extinction. The implications of an observed increase in bar pressing rate during acquisition, well after force stabilization had occurred, were examined.

References

Antonitis, J. J. Response variability in the white rat during conditioning, extinction, and reconditioning. *J. exp. Psychol.*, 1951, **42**, 273–281.

Goldberg, I. Relations of response variability in conditioning and extinction. Unpublished doctoral dissertation, Columbia Univer., 1959.

Schoenfeld, W. N. On the difference in resistance to extinction following regular and periodic reinforcement. Unpublished manuscript. Columbia Univer., 1950.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

Trotter, J. R. The physical properties of bar pressing behaviour and the problem of reactive inhibition. *Quart. J. exp. Psychol.*, 1956, **8**, 97–106.

(Early publication received
April 17, 1959)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

STIMULUS GENERALIZATION AND DISCRIMINATION LEARNING BY PRIMATES[1]

J. M. WARREN AND K. H. BROOKSHIRE[2]

Stanford University

Lashley and Wade (1946) found that reinforcing a single stimulus pattern 200 times did not produce differential transfer effects in subsequent discrimination learning involving the previously rewarded stimulus and a new, neutral stimulus. Experimentally naive chimpanzees and spider monkeys failed to learn discriminations in which the rewarded stimulus was correct more rapidly than discriminations in which the previously rewarded stimulus was incorrect.

These findings are not universally valid. Other workers have obtained contrary data, supporting continuity theories of learning, from experimentally sophisticated monkeys (Grandine & Harlow, 1948; Sinha, 1958; Warren & Hall, 1956). Although such results restrict the generality of Lashley and Wade's (1946, p. 335 footnote) observations, they are not directly relevant to their basic thesis, that transfer does not occur as a result of reinforced responses to a single stimulus by a naive animal, untutored in comparison behavior.

The present experiment was undertaken to provide a more definitive test of the reality of Pavlovian generalization by experimentally naive primates. Lashley and Wade studied only two monkeys and four chimpanzees. Twenty *S*s were used in this experiment.

Because naive monkeys learn pattern discriminations more slowly than object discriminations (Blazek & Harlow, 1955; Harlow, 1945), separate groups were trained with the two sorts of stimuli to determine whether stimulus discriminability affects the magnitude of generalization and transfer.

[1] This research was supported by Grant M-1895 from the National Institute of Mental Health, U. S. Public Health Service.

[2] Now at the University of New Brunswick.

METHOD

Subjects.—Of the 20 macaque monkeys studied, 19 were rhesus, and one was a Philippine. The rhesus monkeys had been tested for 11 weeks on 5 different tests for lateral preferences; the Philippine monkey had been a laboratory pet for 5 yr. All of the *S*s were extremely tame and well adapted to the training situation at the beginning of the experiment, but none had ever been tested on a discrimination task.

Apparatus.—The monkeys were tested in the Wisconsin General Test Apparatus (Harlow, 1949, Fig. 1). The stimuli were presented on a gray tray containing two food wells 8 in. apart.

All of the stimuli used in the experiment were wooden figures, 40 sq. cm. in area. A blue cross and a yellow circle were used in the first replication, and a black square and red triangle in the second replication. The stimuli were presented either as "objects," painted forms 2 cm. thick, or as "patterns." The patterns were made by mounting thin ($\frac{1}{4}$ cm. thick) forms, identical in area and color with the objects, upon white squares 10×10 cm. and 2 cm. thick. Similar stimuli, which are no more easily discriminated than bidimensional patterns, are pictured by Harlow (1945).

Procedure.—The *S*s were randomly assigned to four groups of five animals each, which were subjected to the experimental treatments summarized in Table 1.

In the first replication, all *S*s displaced the blue cross (Stimulus A) 200 times to obtain a preferred food. Only Stimulus A was presented during this pretraining phase, alternating from right to left according to a balanced irregular sequence; Stimulus A was an object for two (O) groups, and a pattern for the other two (P) groups. Discrimination

TABLE 1

EXPERIMENTAL DESIGN

Group	N	Replication I		Replication II	
		Pretraining	Discrimination Training	Pretraining	Discrimination Training
1	5	A+Object	A+B−Object	C+Pattern	C−D+Pattern
2	5	A+Object	A−B+Object	C+Pattern	C+D−Pattern
3	5	A+Pattern	A+B−Pattern	C+Object	C−D+Object
4	5	A+Pattern	A−B+Pattern	C+Object	C+D−Object

learning was begun on the day following completion of pretraining, and the P and O groups were both subdivided into groups for which Stimulus A was correct and Stimulus B (yellow circle) incorrect, and the complementary groups in which B was correct and A incorrect. In terms of expected transfer effects, these groups may be identified by + and −, respectively.

The same general procedure was followed in the second replication. Stimulus C, a red triangle, was presented alone and rewarded 200 times before being paired in discrimination learning with Stimulus D, a black square. Table 1 shows that, for all groups, both the cue and transfer conditions were systematically reversed from the first to second replications. Thus, for example, Group 1 was shifted from the O+ to the P− condition.

The monkeys were given 50 trials per day during both pretraining and discrimination learning phases. The criterion of mastery on the discriminations was 20 consecutive errorless trials, the same criterion used by Lashley and Wade. The noncorrection method was used in discrimination training. Testing was continued for 50 trials on both problems, even if *S* attained criterion in fewer trials.

RESULTS

The results are presented in Table 2, which shows mean errors to criterion as a function of replication, cue, and transfer conditions. These results, quite consistent in the two replications, indicate that the − groups, under both cue conditions, make approximately twice as many errors as the + groups. It is readily apparent that *S*s tested with patterns make four to five times as many errors as *S*s tested with objects under the same transfer conditions.

The distribution of individual error scores was distinctly non-normal, owing to considerable positive skewing in the scores made on the pattern discriminations. For example, two *S*s in the − pattern groups made almost three times as many errors as the group means. Nonparametric significance tests seemed, therefore, most appropriate for evaluating these data. The Mann-Whitney test indicates that the differences between scores on the + and − object discriminations were significant on both the first ($U = 3$, $P = .028$) and second ($U = 2$, $P = .016$) replications. Scores on the − pattern problems were also significantly greater than on the + pattern problems, the values of U being 3 ($P = .028$) and 3.5 ($P < .048$) on the first and second replications, respectively.

No test of the significance of the differences between pattern and object

TABLE 2

MEAN ERRORS TO CRITERION AS A FUNCTION OF TRANSFER AND STIMULUS CONDITIONS

Condition	Experiment 1	Experiment 2
+Objects	6.0	3.0
−Objects	14.8	14.2
+Patterns	31.6	16.8
−Patterns	58.8	62.6

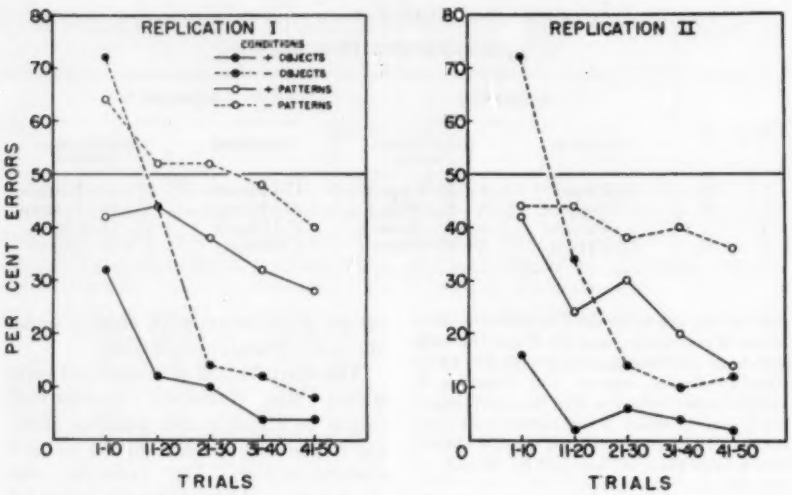


FIG. 1. Intraproblem learning as a function of transfer and cue conditions.

discriminations was necessary. In only one of 20 cases did an *S* in a pattern group make fewer errors on a problem than any *S* in the corresponding object group.

Additional data regarding the nature of the relations between cue and transfer conditions and discrimination performance are presented in Fig. 1, in which intraproblem learning on the first and second replications is plotted for each cue and transfer condition. The intraproblem learning curves for the two replications are highly similar, in spite of more rapid learning of the second task. More importantly, Fig. 1 indicates that reinforced presentations of a single object produce far greater transfer effects than presentations of a single pattern, and that object discriminations are learned much more rapidly than pattern discriminations. Negative transfer effects were largely overcome within 30 trials on the object problems, but large differences persisted between the + and − pattern groups throughout the first 50 trials.

DISCUSSION

The present experiment indicates that experimentally naive monkeys definitely develop adient tendencies toward patterns and objects as a result of reinforced responses to a single stimulus. The monkeys had no opportunity to develop comparison behavior in formal discrimination learning experiments before participating in this investigation. In a personal communication to Grandine and Harlow (1948, p. 335) Lashley has specified this condition as critical for a definitive test of his theory of discrimination learning. It must be concluded, then, that the results reported here are incompatible with Lashley's theory but support continuity theories of learning.

There are two plausible explanations of Lashley and Wade's failure to obtain positive results with respect to the effects of reinforcing a single stimulus on subsequent discrimination learning. It has been pointed out that error scores, and also the variability of the scores, were much greater on pattern than on object discriminations. Lashley and Wade tested but few animals, and only with patterns. Relatively high variability and a small number of *S*s would create

conditions very favorable to sampling errors, and consequent failure to obtain significant differences.

The monkeys studied in this experiment were quite tame at the beginning of the experiment. There is no indication that Lashley and Wade's primates had been tamed as thoroughly. On the contrary, one of their four chimpanzees became emotionally disturbed and did not complete the experiment. It is well known that *S*'s adjustment to the experimental situation is an important determinant of his behavior, particularly on difficult learning tasks. Differential taming is, then, an alternative explanation of the discrepancy between the results obtained by Lashley and Wade and those obtained in the present study.

Summary

Twenty experimentally naive macaque monkeys made 200 reinforced responses to a single pattern or object stimulus. Subsequently, they were required to discriminate between the rewarded stimulus and a new neutral stimulus. For half of the animals, the previously rewarded stimulus was correct, and for the other half, the new stimulus was correct, in discrimination learning.

Two replications yielded essentially identical results:

1. Significant differential transfer effects were obtained; fewer errors were made by *Ss* required to choose the pattern or object rewarded in single presentations.

2. The *Ss* learned to discriminate objects much more rapidly than patterns.

The results of this experiment do not confirm the findings of Lashley and Wade, and are interpreted as a critical negative test of Lashley's theory of generalization and transfer.

References

Blazek, N. C., & Harlow, H. F. Persistence of performance differences on discriminations of varying difficulty. *J. comp. physiol. Psychol.*, 1955, **48**, 86–89.

Grandine, L., & Harlow, H. F. Generalization of the characteristics of a single learned stimulus by monkeys. *J. comp. physiol. Psychol.*, 1948, **41**, 327–338.

Harlow, H. F. Studies in discrimination learning by monkeys: V. Initial performance by experimentally naive monkeys on stimulus-object and pattern discriminations. *J. gen. Psychol.*, 1945, **33**, 3–10.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–56.

Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

Sinha, M. M. Effects of reinforcing single stimuli upon subsequent discrimination learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1958, **51**, 87–91.

Warren, J. M., & Hall, J. F. Discrimination of visual patterns as a function of motivation and frequency of reinforcement. *J. genet. Psychol.*, 1956, **88**, 245–249.

(Received September 26, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

INFLUENCE OF WORK DISTRIBUTION UPON COMPLEX LEARNING BY THE NONCORRECTION AND MODIFIED-CORRECTION METHODS[1]

CLYDE E. NOBLE AND ANTHONY TAYLOR[2]

Montana State University

One of the objectives of our research program on human trial-and-error learning (C. E. Noble, 1957a) is to find the laws connecting such variables as work distribution (W), task factors (T), and amount of practice (N). Functionally speaking, these inter-relationships take the form of equations of the type $R_p = f(W, T, N)$ where $f(N)$ is of the double-exponential family, in which W and T appear as rate parameters (C. E. Noble, 1957b; C. E. Noble, J. L. Noble, & Alcock, 1958). As illustrated in a recent study by C. E. Noble, Alcock, & J. L. Noble (1958), the choice of specific experimental variables for investigation depends on both theoretical and empirical considerations.

Given a sample of *S*s practicing a simple, discrete S-R learning problem (e.g., classical conditioning) with different amounts of rest interpolated either between stimuli (intraserial distribution) or between trials (interserial distribution), Hull's (1943) inhibition theory implies that the more distributed groups would perform with greater proficiency than the relatively massed groups.[3] A great deal of evidence collected on a variety of tasks supports this hypothesis (Hilgard & Marquis, 1940; McGeoch & Irion, 1952; Spence & Norris, 1950). Theoretically, however, whether the effect of work distribution (W) on performance is negative, zero, or positive may be contingent upon task factors (T) like response availability (C. E. Noble, 1955), number of choice points (C. E. Noble, 1957b), or type of choice procedure (C. E. Noble, Alcock, & J. L. Noble, 1958), as well as amount of practice (N). Methodological desiderata suggest further that the task be *E*-paced (C. E. Noble & J. L. Noble, 1958), and that each response entail a constant magnitude of work.

The reason for these qualifications is that reactive inhibition supposedly produces decrements in whatever responses occur in a learning situation. Hence, if the majority of *S*'s responses are incorrect (R−), as in the early phases of a difficult multiple-choice problem, massing of responses should accelerate rather than retard acquisition. Conversely, assuming the same complex task but *S* observed at a later stage of training, continued massing should lead to a relatively slower gain in proficiency than before because his responses would then be predominantly correct (R+). Similar predictions might be made for a

[1] This research was supported by the National Science Foundation under Research Grants G3320 and G4895. Data analyses were performed with the assistance of V. J. Cieutat, J. E. Fuchs, D. R. Gannon, Janet L. Noble, J. C. Roitzsch, and H. A. Walters. An abridged report was made to the Rocky Mountain Psychological Association, May 1959. Helpful comments were received from R. B. Ammons.

[2] Now at Southwestern Medical School, Dallas, Texas.

[3] Because of the complications of treating reactive inhibition (I_R) as a negative drive state, only Hull's (1943) basic notion of I_R as a temporary work-decrement factor will be assumed here.

differential influence of practice schedules upon slow and fast learners, the former being more error prone than the latter.

From the foregoing theory it follows that the *initial* period of compound trial-and-error learning, during which the probability of competing responses is maximal, should benefit from *massed* practice but that the *final* period should benefit from *spaced* practice. In the language of the analysis of variance, we would expect to find a W × N interaction as the acquisition curves of groups training under different distribution schedules first cross, then rise with unequal rates to mastery. Scattered evidence bearing on this deduction may be found in the human learning literature, such as the studies of Cook (1934; 1937), Garrett (1940), and Ericksen (1942) on problem-solving tasks. Other reports reviewed by McGeoch and Irion (1952) suggest null results from the manipulation of W but, as Riley (1952) has pointed out, there are wide methodological differences among these experiments.

The third factor mentioned above, T, remains to be considered. In a previous experiment by C. E. Noble, Alcock, and J. L. Noble (1958), *Ss* practiced an *E*-paced sequence of 10 pushbutton responses on the Selective Mathometer for 20 trials (N) under different types of choice procedure (T). As predicted from generalization concepts, the noncorrection method (one choice per stimulus) produced greater average proficiency and a faster rate of approach to mastery than the modified-correction method (second guessing allowed). Thus, there were significant effects associated with T, N, and the T × N interaction.

Now if the T factor were introduced into the present theoretical context via a similar experiment, one could determine whether type of choice procedure (T) also interacts with work distribution (W). Assuming that C. E. Noble, Alcock, and J. L. Noble (1958) were correct in attributing the inferiority of the modified-correction groups in the earlier study to greater error generalization arising from *Ss*' second-guessing behavior, it seems reasonable to believe that distributed practice would permit more of these competing incorrect responses (R−) to recover from extinction than would massed practice. The resulting heightened interference with the to-be-learned acts (R+) should maximize the T effect under temporal spacing. Presumably, such a differential effect would be reflected in a W × T interaction. On the contrary, if their hypothesis were wrong, and if work decrement (arising from a higher rate of responding) causes the modified-correction method to be inferior, just the opposite prediction would be made. The T effect in that case should be minimized under spaced practice (C. E. Noble, Alcock, & J. L. Noble, 1958, p. 300).

The present experiment is an attempt to resolve this issue as well as to confirm and extend previous findings.

Method

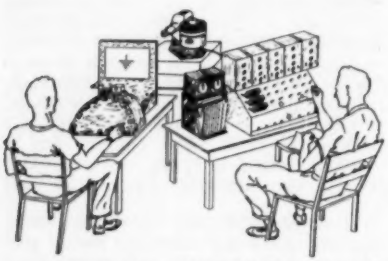
Apparatus.—The learning device, described elsewhere (C. E. Noble & Farese, 1955), was a 19-key serial multiple-choice apparatus called the Selective Mathometer, now fully automatic.[4] A sketch is shown in Fig. 1. Pacing stimuli were presented serially by an electronically controlled slide projector, to which *S* responded by selecting keys on a semicircular panel. The problem was a standard task (C. E. Noble, 1957b) in which 10 keys were available (N_R) and all were

[4] Pending publication, schematic wiring diagrams of the new automatic reward programmer, control panels, and relay circuits for the Mathometer may be obtained at cost by communicating with C. E. Noble.

relevant at the 10 choice points (N_P); thus the ratio $N_P/N_R = 10/10$. The slides consisted of an invariant sequence of 10 circuit symbols, and the correct solution for all *Ss* was to press the odd-numbered keys in the order 15, 7, 11, 1, 5, 19, 9, 17, 13, 3. Automatic (secondary) reinforcement for each correct response (R+) was provided by the onset of a 3-w. green lamp on *S*'s reaction panel, and scores were recorded by polygraph and counters. The time constants for stimulus duration (2 sec.), reward delay (0 sec.), and reward duration ($\frac{3}{4}$ sec.) were the same as in most previous experiments, but both intraserial and interserial intervals differed for the massed and distributed groups. The interstimulus period was 2 sec. for the former and 10 sec. for the latter, while the intertrial periods were 8 sec. and 24 sec., respectively. All data were collected in an air-conditioned laboratory.

Subjects.—The *Ss* were 40 undergraduate students enrolled in the 1957 summer session at Louisiana State University. There were 29 men and 11 women whose ages ranged from 17 to 24 (mean = 18.4 yr.). None had previous acquaintance with the Mathometer.

Procedure.—The *Ss* were divided into four experimental groups of 10 each. They were assigned randomly to the four treatment combinations formed by the two conditions of the W factor (massed vs. distributed) and the two conditions of the T factor (noncorrection vs. modified-correction). Following standard instructions which explained the general problem (C. E. Noble, Alcock, & Farese, 1958, p. 255), the two latter groups received specific directions regarding type of choice procedure (C. E. Noble, Alcock, & J. L. Noble, 1958, pp. 297–298). Essentially, the general instructions emphasized accuracy, quick responding, and the dependency of the green light on correct choices. In the *E*-paced *noncorrection* method *S* is permitted only one reaction at each choice point in the series, whereas in the *E*-paced *modified-correction* method he is free to make alternate responses (second guesses) following an error so long as the slide remains on the screen. All *Ss* received 15 practice trials (N), where N was defined as the number of trials (series of 10 slides) on which reinforcement was *possible*.[5] All *Ss* received 15 trials.



[Printed with permission of *Psychol. Rep.*, 1957]

FIG. 1. The Selective Mathometer.

TABLE 1

MEAN PERCENTAGE OF INITIAL CORRECT RESPONSES (1R+) DURING 15 TRIALS FOR FOUR EXPERIMENTAL GROUPS

Type of Choice Procedure (T)	Work Distribution (W)		Mean
	Massed	Distributed	
Noncorrection	26.20	52.33	39.27
Modified-correction	24.67	43.07	33.87
Mean	25.43	47.70	

RESULTS AND DISCUSSION

Table 1 presents an analysis in percentage terms of the average performance of each group on the problem. The 1R+ measure is used in order to make the two T conditions comparable (C. E. Noble, Alcock, & J. L. Noble, 1958, p. 298). Examination of the cell means indicates that distributed practice results in higher proficiency than massed practice under both T conditions, and that the noncorrection method is superior to the modified-correction method under both W conditions. As expected, the difference due to type of choice procedure is larger under distributed practice than under massed practice.

[5] As pointed out earlier (C. E. Noble, Alcock, & J. L. Noble, 1958, p. 300), the opposite prediction would be made about the relative performance of noncorrection groups if the *S*-paced *correction* method (e.g., maze) were under comparison, for in that procedure an R+ (and hence reinforcement) is guaranteed at every choice point.

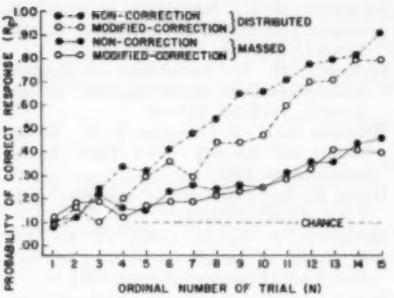


FIG. 2. Compound trial-and-error learning (R_p) as a function of practice (N) under joint variation of work distribution (W) and type of choice procedure (T).

However, when the column means are compared with the row means, a greater main effect is indicated for the W factor than for the T factor.

Figure 2 presents the acquisition curves of the four experimental groups. Probability of making a correct initial choice (R_p) is plotted as a function of amount of practice (N), with work distribution (W) and type of choice procedure (T) as the parameters. The following observations are of interest: (*a*) during Trials 1–3 the massed groups are superior ($\bar{R}\% = 15.7$) on the average to the distributed groups ($\bar{R}\% = 13.3$); (*b*) from Trial 4 onward, however, these positions are reversed, with the two distributed groups rising at faster rates than their massed counterparts; (*c*) the noncorrection groups are generally superior to the modified-correction groups after Trial 3, but the percentage differences are more striking under distributed ($\bar{R}\% = 9.3$) than under massed practice ($\bar{R}\% = 1.5$). All of these phenomena are in agreement with prediction.

To evaluate statistically the various relationships in Fig. 2, a 2 × 2 × 15 (W × T × N) mixed-factorial analysis of variance was performed. The summary shown in Table 2 is based on the number of initial correct responses (1R+) made by the four groups on each of the 15 trials. The main effects of W, N, and the W × N interaction are significant at the .1% level, but none of the other sources of variance reach the conventional 5% level. The numerical values of the T effect and the T × N interaction not only lie in the expected direction (cf. Table 1) but also confirm the preceding experiment (C. E. Noble, Alcock, & J. L. Noble, 1958); however, their *F* ratios are not significant. The weaker of the two ratios, that of 1.63 for the superiority of the noncorrection *Ss* ($df = 1/36$), is equivalent to a *t* of 1.28 ($df = 38$) but the probability that this difference is due to chance is about .20 (two-tailed test).

The graphical evidence in Fig. 2 is suggestive of a greater T effect under spacing, but no statistical basis exists for a W × T interaction ($F < 1.00$). Our differential-error-recovery hypothesis is thereby unconfirmed, although the generalization

TABLE 2

ANALYSIS OF VARIANCE OF INITIAL CORRECT RESPONSES (1R+) DURING TRIALS 1–15

Source	*df*	*MS*	*F*
Between *Ss*	39		
Work Distribution (W)	1	743.70	27.59**
Type of Choice Procedure (T)	1	43.93	1.63
W × T Interaction	1	22.43	<1.00
Error (*b*)	36	26.95	
Within *Ss*	560		
Trials (N)	14	122.48	78.51**
W × N Interaction	14	25.18	16.14**
T × N Interaction	14	2.50	1.60*
W × T × N Interaction	14	1.07	<1.00
Error (*w*)	504	1.56	
Total	599		

* $P < .10$.
** $P < .001$.

explanation of the T effect offered earlier by C. E. Noble, Alcock, and J. L. Noble (1958) still appears tenable. Since the T effect was not *minimized* under distributed practice, the present data fail to support the rival hypothesis that work decrement accounts for the relative inferiority of the modified-correction procedure in paced, complex tasks.

Summary

Serial trial-and-error learning was studied under joint variation of work distribution (W) and type of choice procedure (T). The W factor compared a distributed-practice schedule (10 sec. between stimuli, 24 sec. between trials) with a massed-practice schedule (2 sec. between stimuli, 8 sec. between trials). The T factor compared the noncorrection method (one choice per stimulus) with the modified-correction method (second guessing permitted). Forty human *Ss*, divided into four experimental groups, practiced a sequence of 10 pushbutton responses for 15 trials (N).

As expected from Hull's theory of inhibition, proficiency was superior under massing early in training but under distribution later. There were significant effects associated with W, N, and the W × N interaction. Confirming earlier research, the noncorrection groups were generally superior to the modified-correction groups, but this failed to produce significant T or T × N effects. Despite a predicted greater T effect under distribution, the W × T interaction was also not significant. Due to the absence of a greater T effect in the massed groups (work-decrement hypothesis), however, the hypothesis of error generalization is still tenable as an explanation for the inferiority of the modified-correction method in paced complex learning.

References

Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, **41**, 330–355.

Cook, T. W. Distribution of practice and size of maze pattern. *Brit. J. Psychol.*, 1937, **27**, 303–312.

Ericksen, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, **31**, 339–345.

Garrett, H. E. Variability in learning under massed and spaced practice. *J. exp. Psychol.*, 1940, **26**, 547–567.

Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (2nd ed.) New York: Longmans, Green, 1952.

Noble, C. E. Compound trial-and-error learning as a function of response availability (N_R). *J. exp. Psychol.*, 1955, **49**, 93–96.

Noble, C. E. Human trial-and-error learning. *Psychol. Rep.*, 1957, **3**, 377–398. (a)

Noble, C. E. The length-difficulty relationship in compound trial-and-error learning. *J. exp. Psychol.*, 1957, **54**, 246–252. (b)

Noble, C. E., Alcock, W. T., & Farese, F. J. Habit reversal under differential instructions in compound trial-and-error learning. *J. Psychol.*, 1958, **46**, 253–264.

Noble, C. E., Alcock, W. T., & Noble, J. L. The interaction of type of choice procedure with amount of practice in trial-and-error learning under two reward conditions. *J. Psychol.*, 1958, **46**, 295–301.

Noble, C. E., & Farese, F. J. An apparatus for research in human selective learning. *J. Psychol.*, 1955, **39**, 475–484.

Noble, C. E., & Noble, J. L. Human trial-and-error learning under joint variation of locus of reward and type of pacing. *J. exp. Psychol.*, 1958, **56**, 103–109.

Noble, C. E., Noble, J. L., & Alcock, W. T. Prediction of individual differences in human trial-and-error learning. *Percept. mot. Skills*, 1958, **8**, 151–172.

Riley, D. A. Rote learning as a function of distribution of practice and the complexity of the situation. *J. exp. Psychol.*, 1952, **43**, 88–95.

Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the inter-trial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.

(Received October 1, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

IMMEDIATE VS. DELAYED SHOCK IN T-MAZE PERFORMANCE [1]

ROBERT K. KNAPP,[2] RICHARD H. KAUSE,[3] AND CHARLES C. PERKINS, JR.

Kent State University

The experiments reported below were designed to test certain implications of the view that delay of reward or punishment as such has no effect on the strength of a choice response when highly distinctive stimuli are present during the delay period. Such a view can be made compatible with the greater effectiveness of immediate than delayed rewards (e.g., Logan, 1952) by assuming that the effect of delay is to alter the pattern of internal stimuli that prevails before and during reward presentation and thus to alter the primary reinforcing properties of the stimulus situation. The view that delay as such has no effect will be designated the hypothesis of 100% transfer (of reinforcing properties to delay stimuli).

EXPERIMENT I

The conditions of Exp. I are so designed that if the joint hypotheses of 100% transfer and the effect of preparatory responses on the reinforcing properties of subsequent stimuli (Perkins, 1955) are correct, *S*s will select immediate punishment over delayed punishment. The basis for this rather surprising prediction may be clarified by an examination of Fig. 1 which is a schematic representation of Exp. I.

At the choice point (CP) *S* must either make response R_I, entering delay box X_{DI} or response R_D, entering delay box X_{DD}. R_I is followed by immediate shock, R_D by shock delayed 30 sec. On every trial, confinement in the delay box is terminated after 45 sec. At this time *E* raises a door providing access to a goal box containing food (X_+). After the food has been consumed *S* is removed from the apparatus (X_C). The only difference between the sequence of external stimuli following the two responses is the temporal locus of the shock.

According to the 100% transfer hypothesis the two responses should eventually occur equally often, unless there is differential reinforcement of the responses as a consequence of differences in the reinforcing properties of internal stimuli. In this case there is good reason to expect such a difference in the reinforcing properties of internal stimuli following R_I and R_D.

The best evidence now available

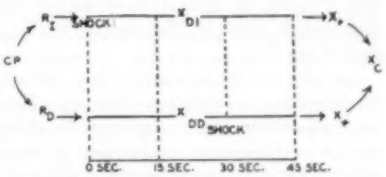


FIG. 1. Schema representing the conditions of Exp. I. The symbols have the following meaning: CP, choice point; R_I, response followed by immediate punishment (shock); R_D, response followed by delayed shock; X_{DD} and X_{DI}, delay boxes; X_+, goal boxes containing food; X_C, removal from the apparatus by *E*. Note that the only difference in the stimuli following R_I and R_D is the temporal locus of shock.

[1] Preparation of this report was facilitated by a National Science Foundation Grant (No. 4839).

[2] Now at Michigan State University.

[3] Now with Goodyear Aircraft Corporation.

(Brogden, 1939) indicates that *Ss* will make conditioned preparatory responses at full or nearly full strength with 50% reinforcement. Thus, conditioned responses that prepare *S* for shock should occur to choice-point stimuli if R_I and R_D occur equally often, as they do in Exp. I.

According to the effect preparatory responses are assumed to have on the level of relative reinforcement of subsequent stimulus situations (Perkins, 1955), these preparatory responses or their stimulus aftereffects increase the reinforcement during shock following R_I and decrease the reinforcement of X_{DD} for at least the first few seconds after R_D.

Since no change in external stimuli occurs during the 30 sec. period between R_D and the delayed shock CRs preparatory to the delayed shock would presumably not be perfectly timed. Thus, they might be too late with the consequence that the delayed shock would be more aversive than the immediate shock (and accompanying preparatory responses) following R_I. Also, if the preparatory responses for delayed shock occurred too soon (e.g., 15 sec. after R_D), they would effect a decrease in the relative primary reinforcing properties of part of X_{DD} (15–30 sec. after R_D) below that of the corresponding period of X_{DI}.

In short, it appears that shock preparatory responses during X_{DI} would become more appropriate than would preparatory responses during X_{DD}. As a result the total level of reinforcement during X_{DI} would be above that of X_{DD}. According to the 100% transfer hypothesis this difference would become effective at the onset of X_{DI} and X_{DD}. This in turn would reinforce R_I more than R_D and the former would occur more frequently on free trials after the preparatory responses had become well established.

Method

Subjects.—The data reported are from 10 naive female albino rats. Five were approximately 200 days old at the start of training, and five, 150 days old. Two additional *Ss* were discarded for failing to meet the criterion of responding described below.

Apparatus.—A single choice-point T maze was employed; all sections of the maze were 5 in. wide and 4.75 in. high (inside measurements). Walls and floor were wooden unless otherwise specified. The entire maze was covered with hardware cloth. The starting box was 10 in. long and was separated from a 20-in. stem section by a guillotine door. The alley perpendicular to the stem section contained the delay and goal boxes. The delay boxes were 17 in. long and had grid floors of .19-in. steel rods .5 in. apart. The guillotine doors at the entrance to the delay boxes were 1.5 in. from the choice point. Similar doors divided delay and goal boxes. Both goal boxes were 18 in. long; the floor of the left goal box was covered with .25 in. wire mesh to provide differential tactile cues. The goal and delay boxes on the left were black, and those on the right, white. The starting box and stem section were medium gray.

Electric shock to the grid floors was 60-cycle AC, 60-v. with about 450,000 ohms resistance in series with *S*. In terms of *S*'s reaction this was a mild but clearly aversive shock. The .7-sec. duration of shock was automatically controlled by means of a photographer's timer. Delay intervals were timed with a stop watch.

Illumination was provided by a 60-w. bulb 3 ft. above the choice point. A black screen was placed between the maze and *E*.

Procedure.—Before preliminary training was initiated, *S* was reduced to 80% of ad libitum body weight by limited feedings every 24 hr. Feedings in individual cages immediately after training periods kept *S* at this weight throughout the experiment.

Five days of preliminary training without shock were given to minimize *S*'s tendencies to hesitate and refuse to enter the delay boxes. On these trials the starting box door and doors to the goal boxes were closed when *S* was placed in the apparatus. The starting box door was opened 15 sec. later. As soon as *S* had passed under the door to a delay compartment this door was closed. The door to the goal box was opened 45 sec. later

providing access to a .4 gm. piece of Purina Laboratory Chow in a food dish at the far end of the goal box. The *S* was returned to the transport cage as soon as the reward had been consumed.

Eight trials a day were given during preliminary training and the subsequent training period. Throughout, water was available during the intertrial interval of approximately 10 min. Whenever four of the day's runs had been made to one side, *S* was forced to the opposite side. Thus, total number of runs to each side was equalized.

The procedure on the 36 days of training differed from preliminary training only in that during training *S* received a shock immediately after entering one delay box and 30 sec. after entering the other. The immediate shock was on the right for half the *S*s, and exactly half the *S*s received immediate shock on the side to which the majority of free runs had been made during the five days of preliminary training.

If *S* did not enter either delay box within 10 min., it was pushed to the choice point. In this manner, *S* was forced to enter one or the other delay box. If *S* had to be pushed on two consecutive free and two consecutive forced trials on two successive days, it was discarded. Two *S*s met this criterion and were discarded. No other *S*s had to be pushed after Day 7.

Five of the *S*s were given eight days of further training. On these days no delayed shock was presented. Otherwise, the procedure was identical to the training series.

Results and Discussion

Figure 2 shows the percentage of free responses to the immediate shock side of the maze on each training day. On the last 10 days combined and on each of Days 34 through 36, every *S* made more free runs to the immediate shock side than to the side on which shock was delayed 30 sec. Since all 10 *S*s would acquire a preference for the same shock condition only 2 times in 1,024 due to chance alone (expansion of the binomial), the preference is significant. It can be concluded that under the conditions of this experiment, a response followed by immediate shock attains greater strength than one followed by delayed shock.

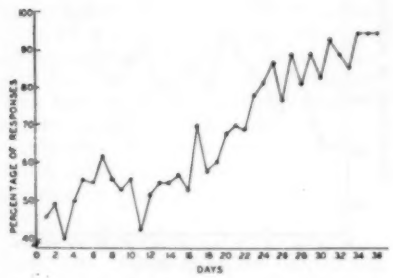


FIG. 2. Results on training trials of Exp. I. The ordinate represents the percentage of free trials to the immediate shock side.

The five *S*s that received further training, during which the delayed shock was omitted, reversed their side preference. The percentages of free runs to the no shock side (previously the delayed shock side) on Days 1–8 of this training were 5, 18, 20, 54, 54, 87, 100, and 100, respectively. Thus, it appears that the earlier acquisition of the response followed by immediate shock could not have resulted entirely from a reinforcing effect of shock or from drive reduction accompanying shock termination.

The results of Exp. I are clearly as predicted from the hypothesis of 100% transfer of relative reinforcing properties and the effect of preparatory responses on level of reinforcement. However, the relative importance of preparatory responses to choice-point stimuli and of the timing of responses when punishment is delayed can not be determined from this experiment as these two factors were confounded.

EXPERIMENT II

Experiment II was designed to determine if precise timing of shock-preparatory responses does increase the relative reinforcing properties of shock or preshock stimuli as our

analysis leads us to expect. The particular design used was intended to maximize the difference in precision of timing of shock-preparatory responses that would occur following two responses that were reinforced by the same sets of external stimuli.

The *S*s were run in a single choice-point maze with a signal sequence followed by shock presented in one arm, and shock followed by signal sequence in the other. The same set of response-shock intervals was used in both arms. The hypothesis that classically conditioned responses in preparation for shock more effectively minimize the aversiveness of the shock if the preparatory responses are precisely timed implies that *S* will acquire a preference for the signal-shock arm.

Method

Subjects.—The *S*s were 10 naive albino rats from the Kent State colony. They were 110 to 155 days old at the beginning of pretraining.

Apparatus.—The apparatus used in Exp. I was modified by the addition of a horizontal row of four flashlight bulbs spaced 1.5 in. apart just over the entrance to each goal box. The frames and the sockets for these lights were painted the colors of the corresponding goal boxes. The bulbs were designated 1 to 4 from left to right (rat's eye view). They were wired in pairs so that whenever one bulb was on, its counterpart in the other arm was also on. The fourth lights (right-hand bulbs) were connected to a buzzer which sounded while these lights were on.

Procedure.—Throughout the experiment *S*s were kept at 80% body weight and given four trials a day with a free-forced technique similar to that used in Exp. I. On all trials *S* was confined to the delay chamber for 3 min. after which he had access to a 0.5 gm. piece of Purina Lab Chow.

No shock or signal lights were presented on the seven pretraining days. During the next 50 days shock (duration and intensity as in Exp. I) was delivered from 30 to 150 sec. after *S* had entered the delay box. The response-shock interval was 30, 45, 60, 75, 90, 105, 120, 135, or 150 sec. Each of these intervals was used equally often in an irregular sequence which was independently determined for each arm.

In the signal-shock arm, onset of Lights 1 to 4 preceded shock presentation by 30, 15, 5, and 2 sec., respectively, and the buzzer accompanied Light 4. The lights were turned off at the onset of shock. In the shock-signal arm the same sequence of lights and buzzer was presented immediately after shock. Thus, the only difference in treatment in the two arms was the sequential relationship between signals and shock.

The signal-shock condition was on the right side for half the *S*s. During pretraining 45.7% of the first two runs of the day were to the arm which became the signal-shock side for that *S*.

If *S* had not entered a delay box within 10 min. of the opening of the starting box door, it was pushed. However, no *S* had to be pushed after the eleventh day of training. On two occasions *E* was unable to run *S*s in one squad of five as scheduled. However, *S*s were kept on their deprivation schedule and skipping a day had no apparent effect on the results.

Results

The percentages of all *S*s' free runs on successive pairs of days that were to the signal-shock side are shown in Fig. 3. Starting about Day 6 a

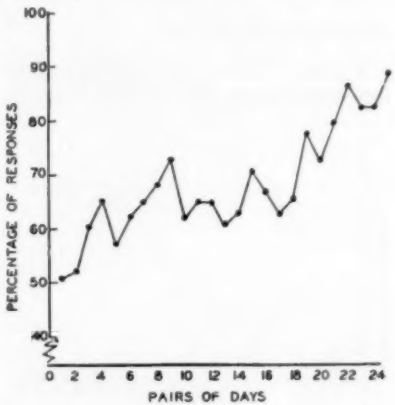


FIG. 3. Results on training trials of Exp. II. The ordinate represents the percentage of free trials to the signal-shock side.

preference for the signal-shock side began to show up. As can be seen in Fig. 3, there was little change in performance between Days 7–36. However, 80–90% of free runs were to the signal-shock side on each pair of days after Pair 21.

Each of the 10 *S*s made more than half of its free responses to the signal-shock side. For every *S*, preference became so marked that during the last five days none had to be forced to the signal-shock arm.

All 10 *S*s would have run to the same condition on the majority of these trials less than 2 times in 1,000 due to chance alone. Thus, the preference for the signal-shock side is significant ($P < .002$), confirming the hypothesis that precisely timed preparatory responses may increase the reinforcing properties (attractiveness) of subsequent stimuli more than poorly timed preparatory responses.

Discussion

The evidence from Exp. I that immediate punishment may be selected over delayed punishment does not prove that delay as such fails to decrease the effects of punishment. Such an effect might occur and be counteracted by the stronger effect of preparatory responses. However, all experimentally demonstrated effects of delay of punishment or reward (e.g., Logan, 1952) on choice responses appear to be consistent with the hypothesis of 100% transfer. This hypothesis has the advantage of greater simplicity than alternative assumptions. Therefore, it is to be preferred unless new evidence requires the assumption that delay as such has an effect.

The results of Exp. I could be interpreted as supporting the hypothesis that *S* will make a response followed by shock if this response also eliminates a danger signal (a stimulus that has regularly been followed by shock), and the accompanying anxiety. However, the results of Exp. II are not in line with such an interpretation, as in this experiment *S*s acquired the response which was followed by a danger signal (lights and buzzer) rather than one which was not.

Caution should be used in generalizing from the results of Exp. II which involved shock to conditions that involve substitution of other aversive stimuli or rewards for shock. In fact, there are reasons for suspecting important differences between the functional characteristics of responses in preparation for food reward and responses in preparation for electric shock. Observation of *S*s suggests that responses in preparation for shock are at least partially postural, striped muscle reactions (e.g., crouching on the grid). One would expect precise timing of such responses to produce a significantly more reinforcing situation than would result if timing were inaccurate. However, responses in preparation for food are presumably primarily glandular (e.g., salivation) and may not require such precise timing.

Results that may be interpreted as supporting this hypothesis have been obtained by Weyant (1957), who failed to find any evidence that a regular signal sequence before a food reward delayed 30 sec after choice increased the reinforcing effect of the food. Apparently the food preparatory responses were equally appropriate with or without the signal sequence. The most likely explanation would seem to be that within limits inexact timing of food preparatory responses does not decrease their appropriateness.

However, since a 30 sec. response-food interval was used on all trials in Weyant's experiment, it is possible that precisely timed preparatory responses occurred even when food was not preceded by the signal sequence. Such timing of preparatory responses might result from internal cues or in Ferster and Skinner's (1957) terms "an accurate internal clock." Further experimentation is clearly needed to learn more about the exact role of responses in preparation for various stimuli.

SUMMARY

In Exp. I, rats were run in a single choice-point maze with food reward delayed 45 sec. on every trial. One response was followed by immediate shock, the other by shock delayed 30 sec. Forced trials were used to equate the number of runs to each arm. All 10 *Ss* acquired a significant preference for the immediate shock.

In Exp. II, rats received food reward delayed 180 sec. on every trial in a single choice-point maze. They received shock between 30 and 150 sec. after either response. On one arm shock was preceded by a 30 sec. signal sequence, on the other the signal sequence followed shock. Number of responses to each arm was equated by forced trials. All 10 *Ss* learned to run predominantly to the signal-shock side on free-choice trials.

It was concluded that all experimental evidence is compatible with the assumption that delay of reward or punishment as such has no effect on choice performance after extensive training when distinctive cues are present during the delay.

REFERENCES

BROGDEN, W. J. The effect of frequency of reinforcement upon the level of conditioning. *J. exp. Psychol.*, 1939, **24**, 419–431.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

LOGAN, F. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, **43**, 393–399.

PERKINS, C. C., JR. The stimulus conditions which follow learned responses. *Psychol. Rev.*, 1955, **62**, 341–348.

WEYANT, R. G. The relation between preparatory responses and the effectiveness of delayed reward. Unpublished master's thesis, Kent State Univer., 1957.

(Received November 10, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

CS TERMINATION AND THE RESPONSE STRENGTH ACQUIRED BY ELEMENTS OF A STIMULUS COMPLEX [1]

DELOS D. WICKENS AND HENRY A. CROSS

Ohio State University

AND

ROBERT M. MORGAN

University of New Mexico

In the learning situation wherein a response is followed by reinforcement, there are many stimuli present in addition to the particular stimulus which is specifically being manipulated by *E*. This is especially true of learning situations outside the laboratory where stringent controls are not placed upon the stimulus environment. The usual statement of reinforcement would seem to imply that the increments in habit strength acquired by these various stimuli as a consequence of the reinforcement would be the same as that acquired by the arbitrarily chosen "cue stimulus."

When this problem was briefly considered by Hull (1943) he concluded that equal habit loadings would not accrue to all the stimuli which were concurrent with reinforcement and he offered several conjectures about various factors which might be responsible for differential growth in response strength. Underlying several of his conjectures was the notion that the variables which influence the development of response strength to a single cue would operate in a like manner for the elements of a complex stimulus.

In a previous study by Wickens (1959) an investigation was made of the effect of time of onset of the different elements of a complex CS with respect to time of onset of the UCS on response strength. The results of this experiment suggested that there was some, but not complete, validity to the hypothesis that the CS-UCS time difference curve obtained from a study of a single stimulus could be used to predict the response strength acquired by elements of a complex stimulus. The present study continues the investigation of the general hypothesis that principles established with the use of single stimuli may be used to predict the response strength of elements of a complex, but a different variable is employed from that of the previous study. More specifically it is concerned with the time of termination of the CS with respect to time of occurrence of the response. The justification for the use of this variable is based principally on the experiments of Mowrer and Lamoreaux (1942) with rats and Wickens and Platt (1954) with humans in which it was shown that in an avoidance conditioning situation acquisition was facilitated if the CS was terminated at the time the response occurred rather than shortly thereafter.

The general procedure of the present experiment consisted of presenting as the CS two independently controlled stimuli, a light and a tone. These came on simultaneously 450

[1] This research was supported in part by a grant, G 2248, from the National Science Foundation.

msec. before the UCS, a shock. One of these was terminated simultaneously with *S*'s response, a finger withdrawal movement, the other persisted for 450 msec. after shock onset.

On particular trials one group of *S*s was tested to the response terminated stimulus alone while the other group was tested to the delayed stimulus alone. Other groups were conditioned to a single stimulus with and without response termination of the CS. A theory which assumed that elements of a complex would acquire response strength in the same manner as stimuli in isolation would predict that any differences found between the two single stimulus groups would also be evidenced in the complex stimulus group on these trials when only one element of the complex was used as the stimulus.

METHOD

Subjects.—A total of 118 men and women students from an introductory course in psychology were used for the final analysis. A few additional *S*s were discarded because of experimental errors.

Apparatus.—The time between onset of the CS and UCS was controlled by an arrangement of wave form generators (Type 162), pulse generators (Type 163) and a power supply (Type 160A) obtained from Tektronix, Inc. The error in this equipment is less than 1 msec. The shock was produced by an AC power supply. One CS was a 1000-cycle tone at approximately 40 db re .0002 dyne/cm.^{2} and produced by a Jackson oscillator and fed into earphones. The other was the illuminating of a 2-w. light behind a ruby panel glass located at eye level and about 6 ft. away directly in front of *S*. The latency of the finger withdrawal response was recorded on a Springfield timer which measured in units of .01 sec. The clock was wired through a microswitch upon which *S*'s finger rested. It was activated by the time control circuit at the inception of the CS and was stopped when *S* removed his finger from the microswitch contact arm. The CS itself was wired through another pole of the microswitch and thus could be terminated by the response. At *E*'s option the microswitch could be bypassed and the duration of CS was then made dependent entirely on the timer. One shock electrode was attached to *S*'s palm and the other was soldered to the microswitch contact arm. Thus when *S* lifted his finger from the electrode he broke the shock and time circuit and also the CS circuit if he was assigned to one of the instantaneous termination single stimulus groups. For the complex stimulus groups one element was always terminated by the response.

The *S*'s arm was restrained by straps around the right forearm just below the elbow and just above the wrist. The microswitch was located along the major axis of the arm with the contact arm and its attached electrode elevated about 1 in. above the surface of the board on which the arm was strapped. The pad of *S*'s middle finger rested on the electrode, and a withdrawal response consisted of an extension of the wrist and finger.

In addition to recording the latency of the withdrawal response, the GSR was also measured by means of the Fels dermohmmeter, serial 117. This instrument permits measurement of *S*'s resistance level at any moment in time. The recording electrodes were attached to the palm and the back of the left hand. An electrode jelly composed of agar agar and zinc sulphate was used between the hand and the zinc electrodes.

The *S* sat in a dental chair in a 6 × 5 ft. semisoundproof room. It was dimly illuminated by a 7.5-w. purple bulb.

An exhaust fan produced a masking noise and together with the earphones and soundproofing prevented *S* from hearing the operating noises of the equipment in the experimental room.

Procedure.—The *S*s were brought into the experimental room, seated in the chair, and given directions. They were told that they were participating in a conditioning experiment; informed about the CS and UCS to be employed and how the UCS could be terminated by lifting the finger from the electrode. In addition they were told that they might acquire a tendency to lift the finger to the CS alone. They were requested not to fight against this tendency or to respond voluntarily.

After the instructions were completed the following procedure was begun: (*a*) Three shock stimulations alone during which the amperage was increased on each succeeding stimulation until a level of 2.5 ma. at 235 v. was reached. This level was used throughout the experiment. This current value was

calculated for an *S* with 50,000 ohms resistance and would accordingly be greater or less for *S*s with resistance below or above 50,000 ohms. (*b*) Two stimulations of the tone or tone plus light for the control and experimental groups, respectively. (*c*) Thirteen training trials, 12 of which were reinforced by shock. Shock was omitted on Trial 8, and in the experimental groups only one of the elements of the complex was presented. (*d*) Five nonreinforced trials, with the experimental groups being tested only to one element of the complex.

The intertrial interval varied by steps of 15 sec. from 30 to 75 sec., with a mean of 54 sec. The same randomized pattern was used for all *S*s. The CS preceded onset of the UCS by 450 msec., and the timer terminated both stimuli 450 msec. later.

The groups.—Four groups of *S*s were run, two with single stimulus CS and two with complex stimulus CS. The single stimulus groups were conditioned to a single stimulus, the tone. For one of these groups the tone was terminated by the response and for the other it was terminated by the timer 450 msec. after shock onset. Its total duration was, in other words, 900 msec. The complex groups were conditioned to the complex of light and tone. These stimuli came on simultaneously, but one was terminated by the response and the other continued until it was cut off by the timer 450 msec. after shock onset. One experimental group was tested on Trial 8 and during extinction to the component which was terminated by the response, while the other group was tested to the delayed termination component. Training was counterbalanced so that the test component was always the tone. There were 30 *S*s in each single stimulus group and one complex stimulus group. The second complex stimulus group consisted of 28 *S*s. This inequality arose from an error in tabulation which was not discovered until some time after the experiment was completed.

Response measures.—Two measures of *S*'s responses were recorded. One of these was the latency (in units of .01 sec.) of the finger withdrawal movement to the time of onset of the CS. For the GSR, the resistance level was obtained at the time the CS was begun, and another reading was obtained of the minimum value of the resistance immediately following the stimulation. Each of these values was converted to a reciprocal and their difference was obtained. The resistance values were read from a meter by *E* and at the sensitivity setting employed, they could be read to units of 200 ohms.

Since a nonparametric method was employed in analyzing the data, no further conversions were performed on either response measure.

TABLE 1

PERCENTAGE OF CRs AND MEAN LATENCY (IN .01 SEC.) BY TRIALS FOR ALL GROUPS

Trial	Single Stimulus				Complex Stimulus			
	Interrupted		Delayed		Interrupted		Delayed	
	% CRs	Latency	% CRs	Latency	% CRs	Latency	% CRs	Latency
2	27	57.5	03	65.7	20	58.9	36	54.1
3	33	53.4	20	59.4	30	53.7	46	48.2
4	23	53.6	30	56.2	47	49.3	43	49.9
5	37	54.9	30	54.8	47	46.0	39	49.5
6	37	50.1	27	54.8	60	44.8	50	47.8
7	47	47.6	30	53.9	40	47.2	46	49.5
8	57	60.6	47	69.9	43	67.2	25	77.6
9	37	50.6	20	56.9	47	46.8	25	56.5
10	50	47.5	40	53.4	37	49.9	50	49.0
11	47	49.5	40	52.8	57	44.3	50	46.4
12	53	45.4	47	51.3	63	43.5	54	47.0
13	43	49.9	47	52.7	73	39.5	57	45.2
14	73	54.8	60	61.9	63	56.8	57	69.0
15	57	58.5	43	68.9	60	61.8	46	69.5
16	60	57.4	57	65.9	50	64.8	43	73.9
17	50	63.2	47	68.3	53	63.6	46	71.9
18	60	57.6	47	71.9	47	69.2	39	71.9

Results

Withdrawal responses.—Data relevant to the performance of the various groups are presented in Table 1, the data consisting of a measure of the frequency of CRs on each trial and the mean latency of the response on each trial. The CRs were defined in the following manner: On trials in which shock was administered, any response of less than 500 msec., while on nonshock trials any response which occurred while the CS sounded. The mean latency of the nonshock trials includes the scores of those who did not respond. This score was 900 msec. No *S* failed to respond during the shock trials.

U tests (Mann & Whitney, 1947) on the latency measures were made between the two single stimulus and the two complex stimulus groups for the trials on which the shock was omitted. The *P* values for each of these comparisons is presented in Table 2. The value is a single-tail value and is based on the hypothesis that the group being tested to the interrupted stimulus would have a shorter latency than the group tested to the delayed stimulus.

In the single stimulus groups there is a rather consistent tendency for the group trained to the response-interrupted stimulus to give shorter latency responses than did the group trained to the delayed-termination stimulus.

A parallel analysis of the performance of the complex stimulus groups suggests a similar trend, with the groups differing at a moderately significant level on Trials 8 and 14, but not for the remainder of the extinction trials.

Galvanic skin responses.—The GSR data were analyzed in the same general manner, and the *P* values for the *U* test on the conductance scores are shown in Table 2. The Interrupted Single Stimulus group shows a consistent tendency to give larger GSR responses than does the Delayed group. Although the trends are in the same direction for the Complex Stimulus groups, a significant level of confidence is reached on only one trial.

TABLE 2

A Table of *P* Values for Avoidance Latency and GSR Magnitude for Nonreinforced Trials

Trial	Single		Complex	
	Avoidance	GSR	Avoidance	GSR
8	.05	.06	.05	.47
14	.09	.01	.02	.07
15	.03	.06	.10	.37
16	.05	.04	.13	.08
17	.15	.03	.08	.03
18	.01	.01	.30	.13

Discussion

The data for the single stimulus groups agree remarkably well with the Wickens and Platt study (1954). In that study, and in the present one, the differences in the frequency of CRs between the two groups was most noticeable in the second trial. In both instances the difference becomes less noticeable with increased training. The present study adds further information concerning the effect of this variable by indicating that the response termination not only facilitates the acquisition of the avoidance response, but also results in a greater magnitude of the GSR. Thus, the results for the single stimulus groups of this experiment support the findings of previous experiments (Mowrer & Lamoreaux, 1942; Wickens & Platt, 1954) and indicate that in a single stimulus situation response termination of the cue stimulus produces superior performance in a conditioning situation.

The question which was raised in this experiment is whether or not this effect will operate upon the stimuli of a complex. The answer seems to be generally in the affirmative. Insofar as the avoidance response is concerned the complex group tested to the response terminated stimulus gave significantly shorter response latencies than did the group tested to longer stimulus on Trial 8, the first test trial, and on the first of the extinction trials. A similar but not significant trend was shown on the other extinction trials.

Further support for the view that it is the experimental variable, rather than chance assignment of better conditioners to this group, which is responsible for this difference can be found by comparing the two groups on other trials. On Trials 2 through 7 the two groups were being treated alike, save for a difference in the modality of the stimulus which was response terminated. As the data on frequency of CRs and latency indicate, there is no consistent tendency for one group to excel the other throughout these trials.

The effect of a favorable element in a complex.—When the performance of the two complex stimulus groups during trials when the complex itself was presented is compared with the Delayed Single Stimulus Group on the same trials, there is a clear tendency for the Complex Groups to excel the Single Stimulus Group. *U* tests were run between the Single Delayed Group and each Complex Group for Trials 2 through 7. In every instance the latency scores favored the Complex Groups. In 7 of these 12 instances the *P* values for the comparison were less than .05 and in 11 instances they were less than .10. This finding suggests that if a complex contains an element favorable to performance and an element less favorable, *Ss* behavior will be controlled primarily by the favorable element. This conclusion is weakened by the fact that there is some tendency for the Complex Groups to excel the Single Instantaneous Group. They are superior at the .10 level or less in 4 out of 11 instances while the Single Group never excels them with a *P* value as low as .10.

There are two plausible interpretations of this trend. One is that the behavior is being determined by two habits, one with the light as the stimulus and one with the tone as its stimulus and these summate to produce response tendencies greater than those occurring to a single stimulus. The other is that a stimulus dynamism effect arises from the energy summation of the two stimuli.

Summary

This experiment tested the hypothesis that a variable which facilitates conditioning to a single CS will function in the same manner upon individual elements of a complex stimulus. The variable employed was the time of termination of the CS with respect to time of occurrence of the response.

The situation employed was avoidance finger conditioning, with the GSR also being measured. A pair of groups of *Ss* was conditioned with a single stimulus—a tone—serving as the CS. For one group the CS was terminated at the time of the withdrawal response and for the other group the CS continued for a short time after the response. The group with the response terminated stimulus showed superior conditioning as measured by both withdrawal responses and GSR.

Another pair of groups was conditioned to a complex stimulus consisting of a light and tone. One of these elements terminated with the response while the other continued for a short time thereafter. On particular trials only one of the elements was presented. For one group this element was the response terminated stimulus and for the other it was the delayed stimulus. Better conditioning on these test trials was evidenced by the group tested to the response terminated stimulus for the withdrawal response but there was little difference between the two groups as measured by the GSR.

It was concluded that the experiment offered limited support for the view that a condition which is favorable for the learning of a response to a single stimulus operates selectively in the same manner upon the elements of a complex stimulus.

REFERENCES

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 52–54.

MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration: A study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54** (5), (Whole No. 247).

WICKENS, D. D., & PLATT, C. E. Response termination of the cue stimulus in classical and instrumental conditioning. *J. exp. Psychol.*, 1954, **47**, 183–186.

WICKENS, D. D. Conditioning to complex stimuli. *Amer. Psychologist*, 1959, **14**, 180–188.

(Received November 17, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

THE ROLE OF PUNISHMENT IN FIGURE-GROUND REORGANIZATION [1]

G. L. MANGAN

Victoria University, Wellington, New Zealand

The research reported in this paper relates to two problems: (*a*) measurement of the degree of impact which given affective states can have upon figure-ground structure; and (*b*) investigation of the conditions governing the direction of figure-ground shifts.

The first problem derives traditionally from Sherif's (1936) original work on the autokinetic phenomenon. It involves consideration of the quantitative relations between the degree of firmness of external structure and the intensity of affect, i.e., amount of reward or punishment. According to Sherif, the extent to which external structure is "falsified," or autism introduced, is a function of these two factors; whether this function is summative or must be expressed in terms of interactions, however, is uncertain. Experimentally, the problem is to vary intensity of affect—the number of reward reinforcements—and measure quantitatively the tendency for figure-ground shifts to occur in a progression of figure-ground patterns varying from those that are as indeterminate as, for example, a Necker cube, to those that have a very steep gradient between figure and ground and are, therefore, difficult to reverse.

The second problem is the traditional one of the effect of reward and punishment in producing response-direction, a problem which has received much attention in learning and perception theory. In studies of figure-ground organization, Schafer and Murphy (1943), Proshansky and Murphy (1942), Smith, Parker, and Robinson (1950) and Smith and Hochberg (1952), using ambiguous material (Schafer-type faces, for example), where the classical determinants of figural dominance—contour, vividness and so on—are minimized, have shown that that aspect of a figure-ground organization which is reward-reinforced will tend to be selected out as figure, and the nonrewarded material structured as ground. This effect has been demonstrated in the auditory (Snyder & Snyder, 1956) and tactual (Sommer & Ayllon, 1956) as well as visual modalities.

Little is known about conditions under which punishing or threatening stimuli dominate competing rewarding or neutral stimuli. Murphy and Solley (1957) consider that two conditions are determinative: (*a*) the opportunity *S* has to escape shock, and (*b*) intensity of shock. Available evidence tends to support this hypothesis. Reece (1954) and Pustell (1954) have shown that unavoidable shock in a noxious situation results in the raising of recognition thresholds for shock-associated percepts, and a number of recent studies (McNamara, Solley, & Long, 1958; Pustell, 1954; Smith & Hochberg, 1952; Sommer & Ayllon, 1956) indicate that a vigi-

[1] This study was supported by a grant from the Committee on Scientific Research, Queen's University, 1957. Thanks are due to Gardner Murphy for advice on design and to R. E. Sanderson and P. Powesland for experimental assistance.

lance-type reaction (the alerting reaction, the lowering of thresholds for punishing stimuli) predominates when shock level is moderate or severe, and that when shock is high, approaching "unbearable," a defense reaction (the raising of thresholds) is more likely. These data, however, are not very extensive, and differences in experimental procedure, stimulus materials, shock levels and so on make interexperiment comparisons difficult. For these reasons, exact weightings of these two factors in producing vigilance or defense reactions are unknown.

In this experiment, an attempt is made (*a*) to establish gradients between degree of firmness of external structure and amount and intensity of punishment in producing figure-ground shifts; and (*b*) to investigate whether punished or neutral material dominates under conditions of mild, severe and beyond-tolerance (bt) shock.

Method

Materials and subjects.—The stimulus material was arranged in four sets. The experimental series consisted of 10 line drawings mounted on white cardboard backing, measuring approximately 12 in. × 12 in. The drawings are shown in Fig. 1.

The 10 test drawings, when arranged in pairs (10 and 1, 9 and 2, 8 and 3, 7 and 4, and 6 and 5) incorporate five degrees of structuredness ranging from marked figural dominance of face or bottle to marked figure-ground indeterminancy. Drawings 5 and 6 are clearly ambiguous, i.e., it is almost impossible to give a judgment of figural dominance for these drawings. This was tested on 15 *Ss*, who were not included in the experimental groups. Without exception, these *Ss* judged Drawings 1, 2, 3, and 4 to be bottles, Drawings 7, 8, 9, and 10 to be faces. In 85% of the cases, Drawings 5 and 6 were judged to be ambiguous.

The remaining three sets of material are denoted conditioning series 1, 2, and 3 (CS_1, CS_2, CS_3). Each of these consisted of 10 pictures, five of which were bottles of various sorts—whisky, deodorant, and so on—the remaining five were profiles or faces, cut from magazines, and mounted on cardboard measuring approximately 30 in. × 18 in.

The *Ss* were 65 volunteer undergraduate men. Thirty of these were arranged in experimental Group 1 (severe shock), with an additional 10 *Ss* as a control group, 10 in Group 2 (mild shock), and 15 in Group 3 (beyond-tolerance shock).

Shock was administered through finger electrodes as a single pulse of 50-msec. duration from a Stoelting stimulator. Shock level for Group 1 was fixed as pain threshold plus half the difference between sensitivity and pain thresholds; for Group 2 as sensitivity threshold plus half the difference between sensitivity and pain thresholds; and for Group 3 as tolerance threshold plus one-third the difference between pain and tolerance thresh-

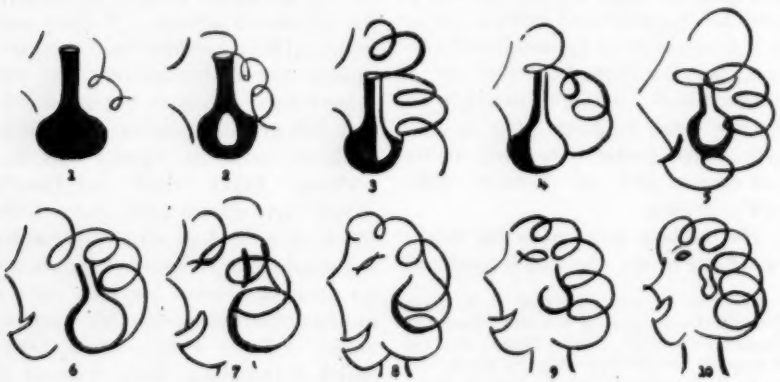


Fig. 1. The 10 drawings employed as stimuli in the experiment.

olds. Sensitivity threshold was defined as a prickling sensation in the two fingers "so that you are only just aware of stimulation"; pain threshold as "the level at which you experience a burning sensation, which you would describe as definitely painful, in your fingers and hand, as if a knife is being pushed up between your two fingers"; tolerance threshold as the point at which pain is extremely severe through the fingers, hand and lower arm—"you feel that it is unbearable, and you want it stopped immediately."

Two thresholds were determined for each *S*, using a modified method of limits. Since only ascending order was utilized, an arbitrary correction for a constant error was taken as the various fractions added to the thresholds. The mean shock intensity for Group 1 was 65 v. ($SD = 7.1$); for Group 2, 36 v. ($SD = 4.8$); and for Group 3, 95 v. ($SD = 8.3$). Shock could not be avoided or alleviated in any way.

Procedure.—The *S* was first of all shown the experimental drawings with the following instructions: "I am going to show you 10 drawings. In some of them you will see two figures, but I want you to name the one which stands out more clearly. Sometimes you may find this extremely difficult or impossible. When you cannot make up your mind which one is dominant, i.e., when the figures appear to balance each other, you may say that the drawing is ambiguous."

The 10 test drawings were held up in random sequence in front of *S*, who was required to make a judgment of figural dominance or ambiguity. Responses were recorded (ES Control).

Electrodes were clipped onto the middle and index fingers of *S*'s left hand, and the two thresholds were determined. The stimulator was placed immediately behind *S*, and operated by an assistant *E*. Onset of shock was thus effectively screened from *S*.

The *S* was then told that he was going to be shown another series of pictures (CS_1), but that this time he was to make no verbal response, merely to look at each picture for the length of time it was exposed. He was warned that sometimes he would be shocked.

The CS_1 pictures in random order were shown to *S*, each for 5 sec., and either the five faces or the five bottles were shocked, each at the end of the first 2 sec. of the 5-sec. exposure. At this point, the assistant *E* moved away from the stimulator and sat beside *E*. The *S* was then reshown the 10 experimental drawings in random order, with instructions as before, and with the assurance that there would be no shock during these presentations. Responses were recorded (ES_1). Two further conditioning series (CS_2 and CS_3) were employed, using the same procedure, and two further sets of responses to the experimental drawings were obtained (ES_2 and ES_3). The order of series presentation in each experiment, therefore, was ES (Control), CS_1, ES_1, CS_2, ES_2, CS_3, ES_3.

There were, in all, 30 presentations of conditioning stimuli, 15 of which were shocked, and 40 judgments of figural dominance, from each *S*.

In the three experimental groups, half the *S*s were shocked for faces, the remainder for bottles. In the control group, five shocks were administered in each of the three conditioning series, but these were randomly distributed between bottles and faces.

RESULTS

Group 1.—In 3 cases out of 30 there were no response changes in the test series. In the remaining 27 a vigilance rather than a defense reaction was shown. The frequency of occurrence of this reaction is highly significant ($z = 5.0$, $P < .00003$) when tested by the sign test. The punished material predominated in all cases. There was an increase in the number of ambiguity responses to bottle cards or in the number of faces reported when faces were being shocked in the conditioning series. Similarly, when bottles were being shocked in the conditioning series, there was an increase in the number of ambiguity responses to face cards or in the number of bottles reported.

The numbers of punished and unpunished figures, ambiguities and figure-ground shifts[2] reported by test series are presented in Table 1.

[2] In the treatment of these data, all response changes in the "right" (vigilance or punished) direction, whether from nonpunished figure to ambiguity response, or from ambiguity to punished figure response, are denoted figure-ground shifts. A weight of plus one (in the case of reversals, minus one) is assigned each such shift.

TABLE 1

RESPONSES IN THE TEST SERIES
GROUP 1

Series	Punished Figures	Ambiguities	Non-punished Figures	Figure-ground Shifts*
ES				
Control	121	51	128	0
ES_1	127	63	110	24
ES_2	129	66	105	31
ES_3	130	81	89	48

* These values are derived in the following manner: For ES_1, for example, $(127 + 63) - (121 + 51) + (127 - 121) = 24$.

A Friedman two-way analysis of variance of the total number of figure-ground shifts occurring shows the effect of treatments, i.e., the number of shock punishments, to be highly significant. The χ_r^2 of 16.5 (2 *df*) is significant at less than .01.

Wilcoxon matched-pairs signed-ranks tests were applied to the differences between all pairs of treatments. All *z*'s except one are significant at better than the .001 level. Regular increase in the size of the coefficients through the matrix strongly indicates a gradient between the number of shifts and number of shock punishments. This is presented in Fig. 2.

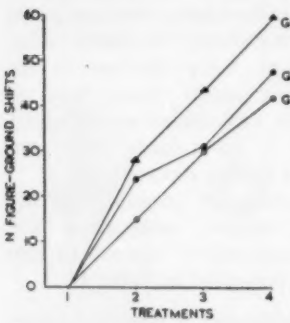


FIG. 2. Increase in number of figure-ground shifts in the three groups as a function of shock conditioning trials. Treatments 1 through 4 are measures obtained after 0, 5, 10, and 15 shock trials, respectively. Frequencies have been corrected for discrepancies in *N* between groups.

For the 10 control *Ss* there were no increases in the number of faces, bottles, or ambiguities reported in any of the test series. The difference between the number of ambiguity responses[3] in ES_3 for control and experimental *Ss*, when tested by a Mann-Whitney *U* test, is highly significant ($z = 2.58$, $P < .005$), with the experimental *Ss* giving significantly more ambiguity responses than the control *Ss*.

Group 2.—In 9 out of 10 cases in the mild shock group, vigilance reactions were shown, i.e., the punished material was emphasized. The frequency of occurrence of this reaction is significant when tested by the sign test ($z = 2.22$, $P < .02$).

A Friedman two-way analysis of variance of the total number of shifts occurring gives a χ_r^2 of 5.85 ($P = .05$). The effect of treatments (the number of shock punishments) is, therefore, significant.

Wilcoxon matched-pairs signed-ranks tests were applied to the differences between all pairs of treatments. As might be expected, under mild shock conditions *z*'s are generally small; the differences between Treatments 1 and 2, and 3 and 4 are not significant. The progression of *z*'s in the matrix, however, follows the expected pattern of increase through the three treatments. The gradient between number of shifts and number of punishments is shown in Fig. 2.

Group 3.—Defense reactions were shown by 13 of the 15 *Ss* who were shocked at beyond-tolerance level. This response-direction is highly sig-

[3] Since no response direction was shown by control *Ss*, increase in the number of ambiguity responses after treatments is the only possible measure of figure-ground shifts.

nificant when tested by the sign test ($z = 2.67$, $P < .004$). A Friedman two-way analysis of variance gives a χ_r^2 of 9.63, which is significant ($P < .02$). As with Groups 1 and 2, the effect of treatments is clearly significant.

Wilcoxon tests between pairs of treatments give z's which are all significant at better than .05. Again, the trend in the matrix is clear and in the expected direction. The gradient is shown in Fig. 2.

Distribution of shifts.—Percentage distribution of shifts for the three experimental groups for all degrees of structuredness are shown in Table 2. With the mild shock group, restructuring initially occurred about equally on the originally ambiguous material of Drawings 5 and 6 and on the next most structured material of Drawings 4 and 7. Under continued punishment, or punishment of greater intensity, however, restructuring shifts to the more highly structured figure-ground patterns. Thus, 50% of Group 2 (mild shock) shifts, 33% of Group 1 (severe shock) shifts and only 23% of Group 3 (beyond-tolerance shock) shifts involve Drawings 5 and 6; with Group 3, 23 of the 30 shifts resulting from the three treatments, and all the 8 shifts occurring in ES_3, are restructurings of Drawings 7 and 4, 8 and 3, and 9 and 2.

TABLE 2

PERCENTAGE DISTRIBUTION OF FIGURE-GROUND SHIFTS FOR THE THREE EXPERIMENTAL GROUPS THROUGH FIVE DEGREES OF STRUCTUREDNESS

Group	Shock	Drawing				
		6, 5	7, 4	8, 3	9, 2	10, 1
2	Mild	50	29.6	21.4	0	0
1	Severe	33.3	48.0	16.7	2.0	0
3	Beyond-tol.	23.3	40.0	23.3	13.3	0

This shift in emphasis is highly significant. The χ^2 derived from the contingency table is 35.59 (6 *df*, $P < .001$). The trend is most clearly shown by considering cell frequencies in Columns 1 and 4 only, the least and most highly structured patterns on which shifts occurred. The 3×2 table gives a χ^2 of 29.57 (2 *df*) which is highly significant ($P < .001$).

DISCUSSION

Gradients have been demonstrated between degree of firmness of external structure and amount of punishment in producing figure-ground shifts where figure-ground patterns were varied from extremely indeterminate to practically irreversible. Judging from the steepness of the gradients and their regularity, the amount of structural falsification induced is very marked, and the relation between the two determining tendencies is probably a simple summative one.

This series of experiments clearly shows that, when shock is unavoidable, one determinant of response-direction is shock intensity. With mild and severe shock, punishment served to emphasize the experimental material to the point where it achieved equal status with the nonpunished, previously dominant material; with beyond-tolerance shock, the response direction was reversed, i.e., the nonpunished percept was emphasized to the point of balance with the physically more dominant punished percept.

These response changes cannot be described strictly as vigilance or defence reactions. The effect is measured more by increase in the number of ambiguities than in the number of punished or nonpunished figures reported. In few cases was a drawing initially judged ambiguous (Drawing 5, for example) identified, after punishment, in terms of the punished or nonpunished figure. Table 2 clearly shows that fewer figure-ground shifts were effected on originally ambiguous material (Drawings 5 and 6) than on the more structured material of Drawings 3 and 4, and 7 and 8. The

process, then, seems to be a selective once, resulting in equal status or balance between punished and nonpunished material, rather than dominance of one over the other.

Presumably, this balance could be achieved either through a process of alienation (or active repression) of the nonpunished material or to a process of emphasis of the punished material when the reaction is one of vigilance, or, when it is defensive, to alienation of the punished or emphasis of the nonpunished material. There is some slight evidence that a process of emphasis is operating. Although this was not originally planned, 10 *Ss* from Group 1, selected on the basis of extremity of reaction, were asked after the experiment to describe, for example, a face in a drawing about which their judgments had changed from "face dominant" to "ambiguous" when bottles were being punished in the conditioning series. In all cases, *Ss* reported that there was little change in the features of the face, but that the bottle now stood out more clearly, judged by the criteria of fixation on the bottle area of the drawing, and of apparent change in the physical features of the bottle—the lines were thicker and blacker, and the area of the bottle relative to the total area of the drawing had increased. At this point, figural balance was achieved, and the drawing was judged ambiguous. Movement appeared to be in the direction of emphasizing the punished, rather than repressing the neutral material.

Within this limited concept of vigilance and defense, these results agree with findings from the studies previously cited (McNamara et al., 1958; Pustell, 1954; Smith & Hochberg, 1952; Sommer & Ayllon, 1956) that mild and severe shock elicit a vigilance reaction, unbearable shock a defense reaction, but contradict those of Reece (1954) and Pustell (1954) that neutral material dominates punished material when shock is inescapable. In this experiment a defense reaction occurred only at beyond-tolerance levels of shock.

There is, of course, an alternative interpretation of the Reece and Pustell data if escape and no-escape conditions are viewed simply as "anxiety-allaying" and "anxiety-eliciting" conditions. Since it is well known that one result of anticipation anxiety is overestimation of intensities of painful stimulation and consequent lowering of thresholds for pain and tolerance, escapability need not be considered as directly a determinant of response-direction, but rather as one factor influencing thresholds, which in turn determine the type of reaction. In these circumstances, there seems little need to implicate directly any factor other than shock intensity in determining the type of reaction.

This, however, is a matter for further research. From the results of this experiment, it is clear that with mild to severe unavoidable punishment one might expect a vigilance rather than a defense reaction, the latter presumably occurring only when the shock is severe enough to force *S* to make some attempt to escape. Tolerance threshold appears to be the critical level beyond which most *Ss* will make such an attempt and show a defense reaction.

Summary

The direction and extent of figure-ground reorganization occurring under punishment conditions were investigated with 55 *Ss* arranged in three experimental groups, each of which was redivided into bottle-shock and face-shock groups. Each *S* gave three sets of judgments of figural dominance to five pairs of drawings, arranged in two sequences, each of which incorporated figure-ground patterns ranging from marked figural dominance of face or bottle to figure-ground indeterminancy, or ambiguity. Thus the drawings, when arranged serially, progressed in 10 stages from bottle dominance to ambiguity to face dominance. Before each series of judgments was taken, *S* was subjected to a conditioning procedure in which 10 pictures were shown in random order, and either the 5 faces or the 5 bottles, which together comprised each series, were shocked. For Group 1 shock level was fixed at above pain threshold, for Group 2 at above sensitivity threshold, for Group 3 at above-tolerance threshold.

The frequency of occurrence of responses in the vigilance direction for Groups 1 and 2, and in the direction of defense for Group 3, were significant, as were the increases in the number of figure-ground shifts occurring after the conditioning treatments. Gradients were shown between number of shifts and amount of punishment. There is some evidence that the vigilance reaction with Group 1 involves emphasis of the punished material rather than alienation of the nonpunished material, and that the process is a selective one, operating in most cases only to the point of figural balance.

References

McNamara, H. J., Solley, C. M., & Long, J. The effects of punishment (electric shock) on perceptual learning. *J. abnorm. soc. Psychol.*, 1958, **57**, 91–98.

Murphy, G., & Solley, C. M. Learning to perceive as we wish to perceive. *Bull. Menninger Clin.*, 1957, **21**, No. 6, November.

Proshansky, H. M., & Murphy, G. The effects of reward and punishment on perception. *J. Psychol.*, 1942, **13**, 295–305.

Pustell, T. E. Cue and drive aspects of anxiety in relation to perceptual vigilance and defense. Unpublished doctoral dissertation, Univer. of Michigan, 1954.

Reece, M. M. The effect of shock on recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, **49**, 165–172.

Schafer, R., & Murphy, G. The role of autism in visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32**, 335–343.

Sherif, M. *The psychology of social norms.* New York: Harper, 1936.

Smith, D. E., & Hochberg, J. The autistic effect of punishment on figure-ground perception. *Amer. Psychologist*, 1952, **7**, 243–244.

Smith, K. R., Parker, G. B., & Robinson, G. A. An exploratory investigation of autistic perception. *Amer. Psychologist*, 1950, **5**, 313–314.

Snyder, F. W., & Snyder, C. W. The effects of monetary rewards and punishment on auditory perception. *J. Psychol.*, 1956, **41**, 177–184.

Sommer, R., & Ayllon, T. Perception and monetary reinforcement: I. The effects of rewards in the tactual modality. *J. Psychol.*, 1956, **42**, 137–141.

(Received November 17, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

A METHODOLOGICAL STUDY OF THE FORM AND LATENCY OF EYELID RESPONSES IN CONDITIONING[1]

KENNETH W. SPENCE AND LEONARD E. ROSS

State University of Iowa

A necessary condition for the discovery of the type of empirical relations or laws the scientist seeks is the identification of the attributes or characteristics of the events being investigated. To this end the psychologist, like every other scientist, must be prepared to specify the end terms (variables) in these relations by means of so-called operational definitions. Such definitions attempt to state the conditions of observation or testing and what must be observed after these conditions have been imposed. One feature of this process of specifying the variables that has not always been sufficiently appreciated is the extent to which trial and error is often involved in it. Frequent redefinition, involving the inclusion of new factors or criteria, the dropping out of old ones, etc., occur in the course of investigation. While this initial hit and miss activity is particularly prevalent in the case of complex behavioral phenomena (e.g., personality traits, social attitudes, etc.), it also not infrequently occurs in connection with simpler types of behavior. One instance of the latter is provided by the manner in which the definition of what constitutes a CR in eyelid conditioning has undergone successive modifications.

In the earliest studies of eyelid conditioning all eye blinks of a given magnitude that occurred between the onset of the CS and the onset of the UCS were identified and counted as CRs. Subsequently, on the basis of a variety of criteria, investigators were able to distinguish two different types of eyelid response (Hilgard, 1931). One class, designated as alpha responses, consisted in a brief closure of the eyelid with a relatively short latency (less than 100 msec.), while another group of eyelid responses involved much more extensive closures, both in amplitude and duration, and considerably longer latencies. Depending upon the CS-UCS interval the latencies of this second type of eyelid response to the CS ranged from approximately 180 msec. upward. It was discovered, moreover, that these two classes of response changed in a very different manner during the course of conditioning. Whereas the alpha responses, which were usually present on the first trial, decreased in frequency of occurrence and amplitude with successive conditioning trials, the second type of response, which was not present initially, increased both in frequency and amplitude. On the basis of these different criteria, including the different lawful relations, *Es* eliminated the alpha responses from the class of responses designated as CRs.

One of the consequences of such a redefinition of this response variable, it should be noted, was the simplification of the law relating the new behavior variable to the practice or conditioning variable, number of paired CS-UCS trials. Under the original definition, two kinds of behavioral events that changed in a quite different, indeed opposite, manner with training were indiscriminately counted together, with the result that a rather complex law

[1] This study was carried out as part of a project concerned with the role of motivation in learning under Contract N9 onr-93802, Project NR 154-107 between the State University of Iowa and the Office of Naval Research.

was being obtained. The redefined response variable led to a simpler law. This is quite characteristic of what happens in the early empirical stage of any science, when the primary job is the discovery of *significant* concepts or variables. That is, such variables must not only be defined for reliability, but they also must be found to have lawful relations with other variables. Simplicity of the resulting laws, it should be noted, is one of the most important criteria determining selection among alternative ways of defining a variable.

Since this original classification of eyelid responses occurring during conditioning into short latency alpha responses and CRs, certain further differentiations among the latter responses have been made. Employing a very intense visual CS, Grant (1945; 1947) obtained evidence for the occurrence of what he termed a beta response to a very bright visual CS. This response, it was found, developed with dark adaptation of *S*'s eye and was entirely independent of the reinforcing UCS. Finding that these light sensitized, beta responses occurred within 200 msec. following the onset of the visual CS, Grant suggested that the complicating effects of such responses could be eliminated by counting as CRs to such visual stimuli only responses beyond 200 msec. While we have not used such intense visual stimuli as Grant in our studies, we have, nevertheless, followed his procedure of counting as CRs to a visual CS only recorded blinks of 1 mm. or more that occur in the interval beginning 200 msec. after the onset of the CS and ending with the onset of the UCS. Thus when the CS-UCS interval is 500 msec., this measuring period is between 200 and 500 msec. following the onset of the CS.

Early in the course of our studies of eyelid conditioning at Iowa, it was observed that some *Ss* gave a high incidence of responses characterized by a sharp, smooth closure of the eyelid, with the eye remaining closed until after the air puff was received (Spence, 1953; Spence & Taylor, 1951). This type of response, it was found, is similar in form to those given by *Ss* instructed to blink to the CS so as to avoid the air puff and by *Ss* who, following conditioning, reported that they were blinking voluntarily in order to avoid the puff. Such patterns of response were observed to occur with especially high frequency when a relatively strong air puff (above 2.0 psi) was employed. Further study of such responses also revealed that they tended to occur predominantly in the interval 200 to 300 msec. The CRs of *Ss* who did not show this response pattern, on the other hand, tended to occur predominantly between 300 to 500 msec. when a 500 msec. CS-UCS interval was employed. In order to eliminate *Ss* who were giving a high incidence of such voluntary responses, we introduced the procedure of not including any *Ss* in our sample who gave 50% or more responses in the interval 200–300 msec. (Spence & Taylor, 1951). This temporal criterion was employed instead of the form characteristic because it was simpler and more objective to use.

The rationale behind this procedure of eliminating *Ss* whose responses were primarily of the short latency, voluntary form was that such responses were governed by different laws than those obtaining in the case of the longer latency class of responses. Thus our studies have indicated that the conditioning performance levels of *Ss* who fall into the voluntary group are not related either to level of a Manifest Anxiety score or to intensity of the UCS as is the case with the nonvoluntary group (Spence, 1953; Spence & Ross, 1957; Spence & Taylor, 1951). Moreover, performance curves of conditioning obtained with such *Ss* differ markedly from those of *Ss* whose responses are predominantly nonvoluntary in form. Not only do they start at a much higher level, but they also tend to reach their asymptotes at a much earlier stage of training.

Since our previous observations of these form-latency relations among eyelid responses during conditioning have been obtained in studies primarily

concerned with other questions, it was deemed advisable to conduct a more systematic investigation of them. The present study represents such an investigation. In it a relatively strong air puff was employed in order to elicit a large number of voluntary type responses. An additional objective was to determine the appropriateness of the 50% criterion that has heretofore been employed as a basis for discarding *S*s as voluntary responders.

Method

Subjects.—The *S*s were 28 men and 58 women students from a summer course in introductory psychology. Two women who gave CRs to test trial presentations of the CS alone were discarded as were two men and two women who misunderstood the instructions. Eighty *S*s remained for conditioning, 54 women and 26 men.

Apparatus and method of recording.—The equipment for recording eyelid closure and presenting the UCS was identical with that used in previous studies (Spence & Taylor, 1951) and will not be described here. The CS was an increase in brightness of a 6-cm. circular disk from .004 to .506 mL. The UCS was a 4.0 psi air puff to the right eye, limited in duration to 50 msec. The CS-UCS interval was 500 msec. with the duration of the CS being 550 msec.

Procedure.—Each *S* was instructed to blink once to the ready signal (the word "ready") and then look at the disk, remaining as relaxed as possible and making no attempt to control his eye. Following the reading of the instructions each *S* received three presentations of the CS alone followed by a single presentation of the UCS. All *S*s were then given 100 conditioning trials. Intertrial intervals of 15, 20, 25 sec., averaging 20 sec. and arranged by a fixed schedule, were used. The interval between the ready signal and CS onset was 2, 3, or 4 sec., randomly varied.

Form judgments.—Two judges, experienced in working with eyelid conditioning records, independently classified each response (on the basis of its form) as being a conditioned response (C), a voluntary response (V), or questionable (?). In terms of combined judgments there were therefore six possible categories for a single response: VV, CC, V?, C?, ??, and VC. Latency measurements were made at a later time by a third person. The criteria used in assigning a response to the V category were (*a*) a sharp, smooth, complete closure, and (*b*) a continuation of the closure until some time after the air puff. An attempt was made to make the judgments as independent of latency as possible.

Any response not meeting the voluntary form criteria was considered to be a CR. While these responses exhibited a wide variety of forms, in contrast to the voluntary responses they were characterized by a gradual, irregular closure rather than a sharp, smooth closure. Moreover, in the majority of instances the eyelid did not remain closed, as in the case of the voluntary responses, but had either opened or was in the process of opening when the air puff arrived.

Results and Discussion

Out of the total of 5684 responses made, 1516 were classified as VVs, 2477 as CCs, 478 as V?s, 775 as C?s, 308 as ??s, and 30 as CVs. Since the number of responses in this last category was quite small it will not be considered further. The latency distributions of the remaining classifications, by intervals of 20 msec., are shown in Fig. 1. It is obvious that the VVs and CCs form overlapping but separate distributions. Approximately 83% of all the VVs made after 200 msec. fall in the 200–300-msec. interval. A feature that should be noted is the relatively large number of VVs in the 180–200-msec. interval, responses that would be eliminated by discarding all responses in the first 200 msec. This is a somewhat higher percentage than we found occurred during this shorter interval in our previous studies, which have employed weaker puffs. Apparently the strong puff in the present experiment tended to reduce somewhat the latency of these voluntary type responses. Actually it was found that of the 196 VV responses in this interval, 188 were made by *S*s who would have been discarded by use of the criterion of 50% responses in the 200–300-msec. interval. In view of the tendency for these voluntary

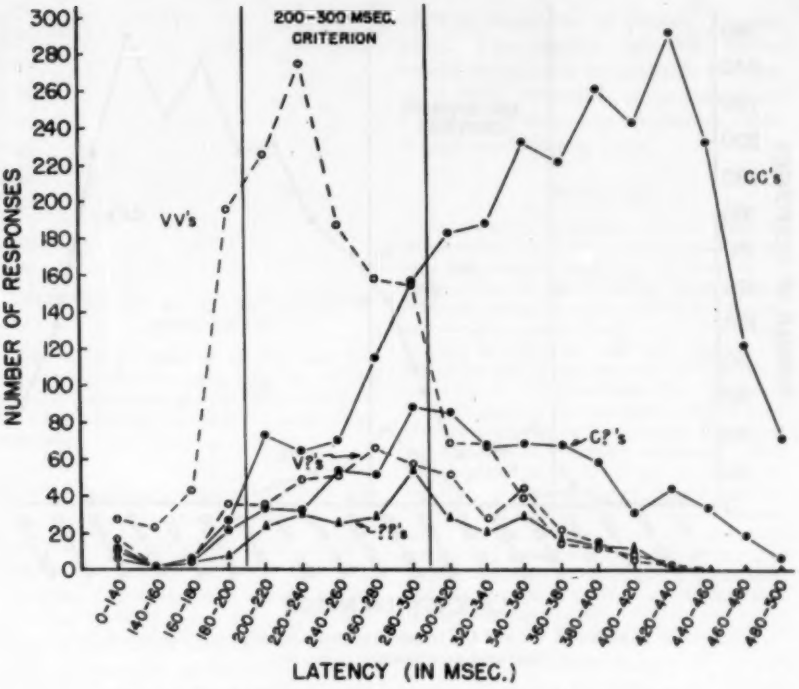


FIG. 1. Total number of responses in each judgment category by successive 20-msec. intervals.

responses to have somewhat shorter latencies with strong air puffs it might be well to extend the criterion interval back to 180 msec. This would especially be the case when auditory and relatively weak visual stimuli are employed as the CS. In such instances there are no beta responses.

When the usual 50% criterion (200–300 msec.) was applied, 31 *Ss*, 6 men and 25 women, were classified as voluntary. With the data of these *Ss* discarded, the latency distributions of VV and CC responses for the remaining 49 *Ss* were as shown in Fig. 2. The criterion is apparently effective in eliminating the majority of voluntary responses from the behavior sample. The percentage response curves, including all categories of responses, for these voluntary and nonvoluntary *Ss* are plotted in Fig. 3.[2] The curves are quite different in a number of respects, the voluntary *Ss*' curve starting at a higher level and reaching a higher asymptote. This asymptotic response level of approximately 90% is the same found for voluntary *Ss* in other studies which employed air puffs of 2.0 psi and below. The present study thus supports the previous finding (Spence & Ross, 1957) that

[2] Because of the large number of VV responses in the interval 180–200 msec. made by some voluntary *Ss* in the present experiment, the percentage response curve for the voluntary group was based on a 180–500-msec. criterion interval instead of that previously employed.

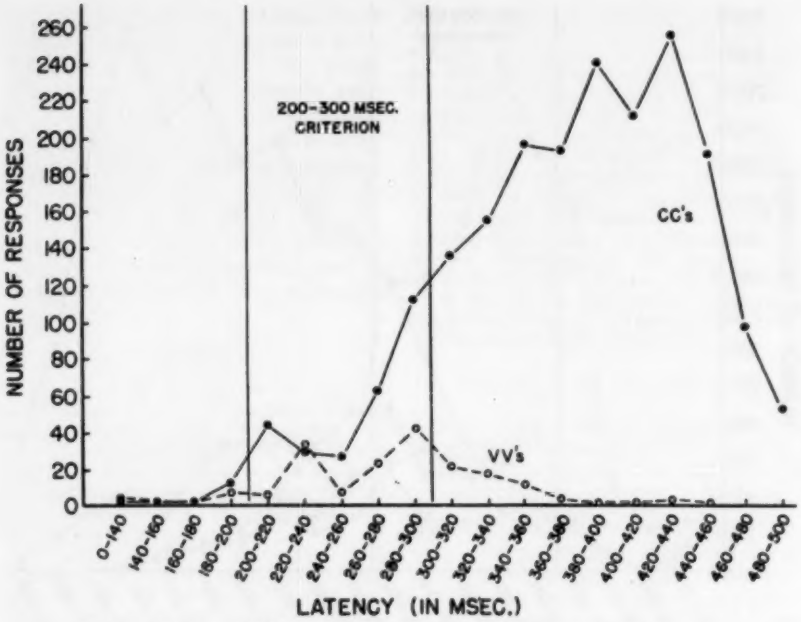


FIG. 2. Number of VV and CC responses remaining after elimination of *Ss* classified as voluntary responders.

the response level of voluntary *Ss* is independent of the strength of the UCS. Another characteristic difference between the two classes of *Ss* is that the voluntary *Ss* show a sudden jump from zero response to maximum response (approximately 90%) whereas the nonvoluntary *Ss* exhibit a gradual increase in percentage of CRs. The gradual increase in the curve of the voluntary group results from different *Ss* beginning to make anticipatory responses at different stages of training.

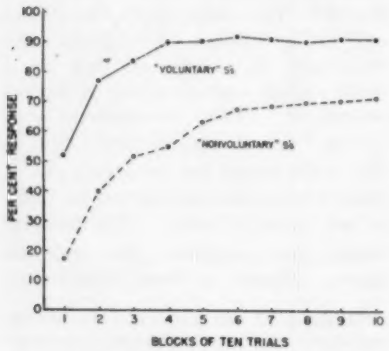


FIG. 3. Mean percentage of response in successive blocks of 10 trials made by voluntary and nonvoluntary *Ss*.

While in past experiments *Ss* have been discarded if 50% of their responses fell in the 200–300-msec. interval, there is a possibility that the use of some other percentage might be more satisfactory in eliminating voluntary form responses from the data. To check this, *Ss* were discarded using criteria from 90% to 10%, in 10% steps. That is, *Ss* were discarded if they had given over

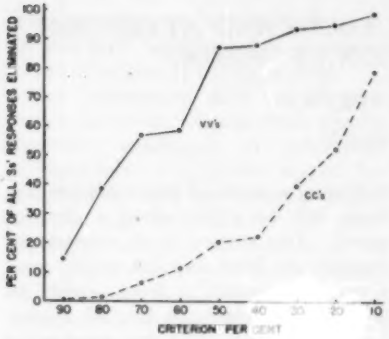


FIG. 4. Percentage of all VV and CC responses eliminated as a function of the criterion (percentage of responses in the 200–300 msec. interval) used to define a voluntary responder.

90% of their responses in the 200–300 msec. interval, over 80% in that interval, etc., to over 10% in that interval. It may be seen that the criterion becomes more severe as we go from a 90% to a 10% criterion, since in this latter case any *S* making 10% or more of his responses in the interval would be dropped from the data. Next, the percentage of the CC and VV responses for all 80 *S*s that would be eliminated by each of these criteria was computed. The results are presented in Fig. 4. It may be noted, for example, that when the 50% criterion was used, discarding the 31 *S*s eliminated 87.6% of the VVs and 20.08% of the CCs. In choosing a percentage criterion a compromise must be reached between eliminating VV responses, which is desired, and discarding CC responses, which is to be avoided. From Fig. 4 it appears that either a 40% or a 50% criterion is the best choice. The use of a more severe criterion would discard very few additional VVs and would eliminate increasingly larger numbers of CCs, while a less severe criterion would allow large numbers of VV responses to remain in the data. The results of this study would thus seem to support the use of a 50% criterion as a means of eliminating voluntary responses from eyelid conditioning data.

SUMMARY

In order to investigate criteria used to define voluntary eyelid responses and voluntary responders in eyelid conditioning, 80 *S*s were given 100 conditioning trials with a relatively strong air puff as the UCS. A large number of responses judged to be voluntary in form were elicited and were found to fall predominantly in the interval 200–300 msec. following the onset of the CS. Dropping from the data *S*s who had given 50% or more of their responses in this interval eliminated the majority of voluntary responses. With the exception of 40%, the use of other percentages as the criterion proved unsatisfactory. The percentage of response acquisition curves of the voluntary and nonvoluntary *S*s were found to differ, the curve for the voluntary *S*s starting at a higher level and reaching the asymptote sooner than that of the nonvoluntary *S*s.

REFERENCES

GRANT, D. A. A sensitized eyelid reaction related to the conditioned eyelid response. *J. exp. Psychol.*, 1945, **35**, 393–402.

GRANT, D. A., & NORRIS, E. B. Eyelid conditioning as influenced by the presence of sensitized beta responses. *J. exp. Psychol.*, 1947, **37**, 423–438.

HILGARD, E. R. Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound. *Psychol. Monogr.*, 1931, **41**(1), (Whole No. 184).

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPENCE, K. W., & ROSS, L. E. Experimental evidence on the relation between performance level in eyelid conditioning and anxiety (drive) level. *Tech. Rep. No. 5*, Office of Naval Res. Contract N9 onr-93802, 1957.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

(Received November 21, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

THE EFFECTS OF RETINAL LOCUS AND ATTENTION ON THE PERCEPTION OF WORDS [1]

HERBERT S. TERRACE

Harvard University

A study by Mishkin and Forgays (1952) which reported a test of equipotentiality in human vision served to stimulate a series of further investigations which have attempted to check the validity of this test (Forgays, 1952; Heron, 1957; Hogg, 1957; Orbach, 1953). Since these and other relevant studies have been reviewed by Heron (1957), and since the experiment to be reported below logically follows from Heron's study, discussion will here be limited to remarks concerning the nature of the problem and an important defect in Heron's design.

Mishkin and Forgays demonstrated that the tachistoscopic presentation of English words either to the right or the left of a fixation point resulted in better recognition in the right visual field. Heron found that there was no difference between right and left recognition scores when either familiar or nonsense geometric forms were used as stimuli. When letters were tachistoscopically exposed, either to the left or to the right of a fixation point, however, more letters were recognized in the right visual field.

Heron interpreted his and previous findings in terms of postulated attentional processes which operated after the tachistoscopic exposure. These attentional processes are assumed to be analogous to the two main types of eye movement that occur while reading English. Carmichael and Dearborn (1947) describe the first kind of eye movement as consisting of a series of short movements from left to right along a line of print. The second type consists of movements from one line to the next ("return sweeps"), from right to left. Thus, when the English reader looks at a line of print, two tendencies control the movement of his eyes. One is to fixate at the beginning of the line, and the other is to move the eye in saccadic jumps from left to right. The tendency to start at the beginning of a line is considered to be the dominant one (Carmichael & Dearborn, 1947).

Heron postulated that when alphabetical stimuli are presented in the right field these two tendencies would operate in the same direction. Beginning at the first letter or word in the right visual field, and continuing along the line, requires only one direction of movement. For material presented in the left field, however, the eyes must move first to the left, to locate the beginning of the line, and then to the right. Thus, under conditions of random successive presentations of alphabetical material the right side would be favored. Since geometrical forms or nonsense figures are not usually read in unidirectional sequences they should be recognized equally easily in both fields.

The results of experiments involving the successive presentation of words or letters in either visual field seem to justify Heron's hypothesis that when alphabetical material is presented, a different postexposural attentional process operates in each

[1] Based on a thesis submitted for the M.A. degree at Cornell University.

field: scanning in two directions in the left field, as opposed to scanning in one direction in the right field.

One important flaw in Heron's design, however, renders these results somewhat equivocal. If differences in right-left recognition scores for alphabetical and geometrical material are to be attributed to postexposural attentional processes, all pre-exposural sets must be controlled. Randomizing the side in which a stimulus will appear does not control for the possibility that *S* might be set for that particular type of stimulus. For example, pilot study records of eye muscle potentials (calibrated to show deflections as small as 2° 30′) indicate that in a series in which only letters are presented, *S* tends to shift his fixation point to the right of the central fixation point. On the other hand, records obtained during a random series of words and forms conclusively show that *S*'s eyes are fixated at the fixation point (Terrace, 1958).

Only after one has controlled for a set for the type of stimulus to be presented can one attribute any difference between the ease of right-left tachistoscopic recognition of words to a postexposural process. This can be accomplished by randomly varying both the type of stimulus to be presented (alphabetical or geometrical) and the side on which a stimulus will appear.

If no differences in recognition scores for alphabetical material are found in an experiment that controls for the set for the type of stimulus to be presented, earlier experiments that did demonstrate such differences (Forgays, 1952; Mishkin & Forgays, 1952; Orbach, 1953) would have to consider this type of set in interpreting their data. If, on the other hand, a difference between right-left recognition scores for alphabetical material is found, one can safely attribute these differences to postexposural processes.

Method

A projection tachistoscope, operated by a Wichita Model 261 electronic timer, with an exposure of approximately 100 msec. was used. The illumination during the exposure was 7.31 ft.-L. (measured with a Luckiesh-Taylor GE brightness meter). This illumination was maintained throughout the experiment by using a second projector to illuminate the screen between exposures. The fixation point, which was a small black x, was on a screen, 9 ft., 10 in. away from *S*'s eyes. A chin rest was used to keep this distance constant and to reduce head movements. All viewing was binocular.

Ninety-six stimuli were used; 16 during a practice session and 80 during the experimental session. Half of the stimuli were words; the other half were unfamiliar nonsense forms. Eight of each type were presented during the practice session and the remaining 40 of each were used as experimental stimuli. Twenty of each type of stimulus appeared to the left of the fixation point; 20 of each type appeared in the right visual field. Thus, there were four experimental conditions: (*a*) a word in the right visual field; (*b*) a word in the left visual field; (*c*) a form in the right visual field, and (*d*) a form in the left visual field. All of the words used were taken from the Thorndike-Lorge list (1944), and were selected from the group of words with a stated frequency of 50 to 100 times per million printed words.

The words and forms were printed in India ink on white typing paper, photographed, and mounted as 35-mm. slides. The words were printed in lower case letters with a Leroy lettering guide. The horizontal angular distance from the fixation point to the midpoint of each stimulus was approximately 4° 40′. Each stimulus subtended a visual angle of approximately 1° 54′. All of the geometric stimuli were between two and three times as tall as the words.

Thirty *S*s were used; 17 men and 13 women. All were college students who could read English fluently. No *S* had any experience with a language that was read from right to left.

The type of stimulus to be presented and the side on which it was presented were randomly varied. When words were used as stimuli, *S* wrote what he saw after each

exposure. When nonsense forms were used, *S* was tested by the recognition method in which three choices were supplied after each exposure. One choice was a copy of the form that was presented; the other two each varied some feature of the exposed form.

To further control for a set for the type of stimuli to be presented, each *S* was trained before the beginning of the experiment to use a card that covered the answer space for a given stimulus until the stimulus had been presented. A typical sequence was as follows: *S* placed his head in the chin rest and informed *E* that he was ready for an exposure. The *E* flashed the stimulus on the screen *after* which *S* moved the card on the answer sheet down to the *top* of the space for the answer to the next stimulus. Adequate markings were made on the answer sheets so that *S* would have no trouble manipulating the card. If a word appeared on the screen, *S* would find a blank space in which to write what he saw. If a form appeared, *S* would find a set of three forms from which he selected the one that he thought appeared on the screen.

All instructions given to each *S* covered the following points: (*a*) the use of the chin rest; (*b*) where to fixate on the screen; (*c*) the tachistoscopic method of presenting stimuli; (*d*) the types of stimuli to be presented (words and forms); (*e*) where the stimuli would appear (to the right or the left of the fixation point); (*f*) the random variation of the type of stimulus and the side on which it would appear; (*g*) the use of the answer booklet and card; (*h*) the use of a push button to signal *E* that *S* was ready for the next stimulus; and (*i*) the sequence of events in each trial.

Each *S* was given 16 practice exposures involving 4 words on the left, 4 words on the right, 4 forms on the left, 4 forms on the right—presented in random order. After the practice session, any further questions on procedure were answered. Rest periods were provided if requested.

TABLE 1

MEAN RECOGNITION SCORES

Stimuli	Side		Diff.	*CR*
	Right	Left		
Words	16.92	12.16	4.76	2.12*
Forms	12.80	14.00	1.20	.75

* For $P = .025$, $CR = 2.045$.

Both the practice and the experimental sessions were conducted in a darkened room. The only illumination came from the constantly illuminated screen and a pair of small light boxes placed next to the table in front of *S* to aid him in writing his answers.

Scoring was based on one point for each correct answer; no partial credit was given. The maximum score possible for each of the four experimental conditions was 20.

RESULTS

The mean recognition scores for words and forms on the right or left side appear in Table 1. *t* tests were used to test the differences between the mean scores for word and form recognition on the right and left sides. The critical ratios for the differences between the two sets of means tested appear in the lower part of Table 1. The .025 level of significance was used to keep the level of error for both comparisons at the .05 level. Inspection of Table 1 shows that the difference between scores for word recognition on the right and left sides was significant while the difference for forms was not. Much more impressive than this finding is the fact that 29 of the 30 *S*s showed higher recognition scores for words on the right side; the recognition scores for forms showed that 17 were higher on the left side, 7 were higher on the right side, and that 6 were ties.

DISCUSSION

The use of nonsense geometric forms to control for pre-exposural tests did not seem to affect the superiority of tachistoscopic recognition of words in the right visual field. It is, therefore, possible to attribute the differences in right-left word recognition scores to postexposural processes.

The first stages of these postulated postexposural processes are probably concerned with discriminating the nature of the stimulus, that is, whether or not it is alphabetical. There is much evi-

dence that the ability to make this discrimination is a function of reading experience. Forgays (1952) found that no differences could be obtained in right-left recognition scores for words in *Ss* below the seventh grade of elementary school. Hogg (1957) has shown that a group of college students were able to recognize more tachistoscopically presented words than a group of soldiers whose formal schooling did not go beyond elementary school training. Solomon and Howes (1951) have reported that the threshold for the recognition of words is a function of the frequency of their prior use. These findings are in agreement with evidence presented by Gibson and Gibson (1955) from which they conclude that the accuracy of discrimination of a stimulus is a function of experience with similar stimuli.

Once the nature of the stimulus has been discriminated, an attentional process, specific to that type of stimulus, can begin to operate. In the case of tachistoscopically presented words, the temporal aspects of this process seem to be governed by the spatial features of the stimulus. This process would last as long as the neural trace of the stimulus remains clear. How this process operates seems to be adequately described by Heron (1957). The first letter or word of the stimulus is attended to. Then the remainder of the stimulus is scanned. For stimuli on the right side of the fixation point only one direction of scanning is necessary; on the left side two directions are necessary, first to the left and then to the right.

Summary

An experiment was conducted to test the presence of any differential recognition effects for tachistoscopically presented words to the right and the left of a centrally located fixation point when the set for the type of stimulus to be presented was controlled by the inclusion of geometric stimuli in the test series. The results of this experiment, which showed that more words were recognized in the right visual field, are in agreement with those of Heron's study in which only the side on which a stimulus appeared was controlled. The role of postexposural attentional processes was discussed as a possible explanation for the differential recognition effects observed.

References

Carmichael, L., & Dearborn, W. H. *Reading and visual fatigue.* Boston: Houghton Mifflin, 1947.

Forgays, D. G. The development of a differential word recognition. *J. exp. Psychol.*, 1952, **45**, 165–168.

Gibson, J. J., & Gibson, E. J. Perceptual learning—differentiation or enrichment? *Psychol. Rev.*, 1955, **62**, 32–40.

Heron, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

Hogg, B. Effects of letter position on recognition. Unpublished master's thesis, McGill Univer., 1957.

Mishkin, M., & Forgays, D. G. Word recognition as a function of retinal locus. *J. exp. Psychol.*, 1952, **43**, 43–48.

Orbach, J. Retinal locus as a factor in recognition of visually perceived words. *Amer. J. Psychol.*, 1953, **65**, 555–562.

Solomon, R. L., & Howes, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

Terrace, H. S. The effects of retinal locus and attention on the perception of words. Unpublished master's thesis, Cornell Univer., 1958.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words,* New York: Teachers' Coll., Columbia Univer., 1944.

(Received December 15, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

FACTORS WHICH INFLUENCE SUBJECTIVE PROBABILITY

S. S. KOMORITA

Vanderbilt University

One of the primary concepts in the area of decision making under uncertainty is psychological, or subjective, probability, i.e., an individual's judgment or estimate of the likelihood of an event. In recent years, there have been a number of studies (Cohen, Dearnaley, & Hansel, 1957; Davidson, Suppes, & Siegel, 1957; Edwards, 1953; Preston & Baratta, 1948) attempting to measure subjective probability. This study is an extension of these studies in an attempt to determine variables which influence subjective probability.

METHOD

Estimating subjective probability.—As a measure of subjective probability, *Ss* were asked to estimate the "odds" against the occurrence of a number of events. This seemed to be a reasonable procedure since there is a 1 to 1 correspondence between the "odds" for or against the occurrence of an event and the probability of that event. Accordingly, the following transformation was used to determine subjective probability, denoted by $\phi(p)$:

$$\phi(p) = \frac{n}{n + n'} \qquad [1]$$

where n and n' denote integers used in quoting "odds" for a fair bet. For example, if *S* responded, "2 to 1" (*against* the occurrence of an event), $n' = 2$, the first integer; $n = 1$, the second integer; and we have, $\phi(p) = 1/(1 + 2) = 1/3$.

The variables.—Three experimental variables were used and will be hereafter referred to as Variables A, B, and P, respectively. Variable A consisted of the number of possible events which could occur in a game. All events consisted in the simultaneous toss of 10, 100, or 500 coins, hereafter referred to as A_1, A_2, and A_3, respectively. The task of *S* was to estimate the odds against a certain number of heads coming up if the coins were to be tossed. Note that with respect to the number of heads which can occur, there are 11, 101, and 501 possible outcomes for A_1, A_2, and A_3, respectively.

Variable B consisted in estimating odds using numbers only vs. using actual monetary values, hereafter referred to as B_1 and B_2, respectively. The hypothesis underlying Variable B was that estimating the odds using monetary values would differ from the use of nonmonetary values; i.e., that the utility of money would influence *S*'s estimate of the odds.

Variable P consisted of four probability values. These values were .17, .38, .62, and .83, hereafter referred to as P_1, P_2, P_3, and P_4, respectively.

Subjects.—The *Ss* were 72 male undergraduate students enrolled in an introductory psychology class and their participation as *Ss* fulfilled part of their requirements for the course.

Experimental design.—Variables A and B were factorially combined so that there were six experimental groups with 12 *Ss* in each group. All *Ss* were asked to make judgments with respect to each of the four probability values, and consequently, repeated measures on Variable P were obtained.

Instructions and procedure.—Groups B_1 were given the following instructions:

"Most of you are probably familiar with the concept of 'odds' used in games of chance. For example, in betting on the outcome of a game, if we say that the odds are 2:1 (2 to 1) against a particular team winning, this means that the team *does not* have a very good chance of winning. On the other hand, if we say that the odds are 1:2 (1 to 2) this means that the team *has* a good chance of winning.

"This is an experiment to determine how accurately you can estimate the odds for the occurrence of various events. For each event presented, your task is to place a number on one side of a colon. If you feel that the event is very *likely* to occur, the number on the *left* should be *smaller than* the number on the *right*; e.g., 2:5. On the other hand, if you feel that the event is *not likely* to occur, the number on the *left* should be *larger than* the number on the *right*; e.g., 10:3. If you feel that the event is just as *likely to occur as not to occur*, then the two numbers should be equal; e.g., 1:1.

"It should be emphasized that this is not an experiment to determine how accurately you can mathematically calculate the odds. Consequently, *do not compute the odds* even if you are able to do so. Merely estimate the odds as accurately as possible on the basis of your personal judgment. Furthermore, once you have made a judgment for an event, do not go back to see what you put down for that event."

Groups B_2 were given essentially the same instructions but with the following modifications:

"For each event presented, you are to assume that you are going to bet someone that the event *will* occur or *will not* occur. Your task is to fill in an amount of money in a blank space on one side of a colon such that the bet will be a *fair one*; i.e., so that you would not care whether you were betting *for* the occurrence or *against* the occurrence."

"Realistically assume or imagine that you are going to bet someone that the event *will* or *will not* occur, and estimate the odds as accurately as possible on the basis of your personal judgment."

In addition, all groups were given the following instructions:

"The following events consist of the simultaneous toss or flip of 10 coins (100 and 500 for other two groups). After each toss, the number of heads is to be recorded. For each of the following events, estimate as accurately as possible what the odds should be by placing a number (or value) in the blank space."

The stimuli.—The stimuli consisted of four probability events each typewritten on a strip of paper and stapled together in random order. A sample stimulus event for Group A_1B_1, e.g., was as follows:

Obtaining at least 5 heads (5 or more). ———:100.

For half of the B_1 groups, the number 100 was inserted on the left of the colon and the task was to insert a number on the right of the colon. For the other half, the number 100 was placed on the right of the colon and the task was to insert a number on the left of the colon.

For half of the B_2 groups, the value of \$10.00 was inserted on the left of the colon, and the task was to insert a monetary value on the right of the colon. For the other half, the value of \$10.00 was placed on the right of the colon, and the task was to insert a monetary value on the left of the colon.

For each of the six groups, therefore, a number 100 or the value \$10.00 was either on the right or the left of the colon, and *S*'s task was merely to insert another number or value in the blank space. It was felt that this would control for the effects of the magnitude of the numbers or monetary values which might be used by *S*s in estimating odds. For example, some *S*s might respond, "\$5.00 to \$1.00," while others might respond, "\$0.05 to \$0.01," and the magnitude of these values might influence their estimate of the odds.

TABLE 1

SUMMARY OF STIMULUS EVENTS PRESENTED TO EACH GROUP

A_1(10 coins)		A_2(100 coins)		A_3(500 coins)	
X	p	X	p	X	p
4	.828	46	.816	240	.826
5	.623	49	.618	247	.623
6	.377	52	.382	254	.377
7	.172	55	.184	261	.174

With respect to Variables A and P, Table 1 represents a summary of the probability stimuli which were presented to *S*s. Each stimulus consisted of a statement of the form, "Obtaining at least X heads (X or more)," and Table 1 shows the value of X presented to each of the A_1, A_2, and A_3 groups as well as the probabilities of these events.

The response measure.—Note in Table 1 that the objective probabilities for the three groups are not identical. These differences are not critical, however, since the response measure used throughout the analysis was absolute deviation of subjective probability from the objective probability. That is, each estimate of odds was transformed to a subjective probability value and this value was subtracted from the objective probability level. The absolute difference was then used as the response measure.

RESULTS AND DISCUSSION

The summary table for the analysis of variance is given in Table 2. Bartlett's test for homogeneity of variance indicated that the variances of the six groups were not significantly different. Table 2 shows that the main effects for Variables A and P are significant at the 1% level, and further, that the interaction between A and P is also significant at the 1%

TABLE 2

SUMMARY TABLE FOR THE ANALYSIS OF VARIANCE

Source	*df*	MS	*F*
A	2	447.362	7.365*
B	1	181.928	2.995
A × B	2	130.843	2.154
Error (B)	66	60.744	
Total	71		
P	3	3782.542	108.572*
A × P	6	262.219	7.527*
B × P	3	88.233	2.533
A × B × P	6	46.870	1.345
Error (W)	198	34.839	
Total	216		
Total	287		

* $P = .01$.
Note.—All measures were multiplied by 100 for convenience of computation.

level. These results indicate that the accuracy of *Ss* in estimating the probability of events seems to vary with the number of events or outcomes which are possible, and also with the probability value of the events. This generalization must be qualified, however, since the interaction between the two variables was also significant.

Figure 1 summarizes the data with respect to Variables A and P and represents a graph of the relation between objective probability, p, and subjective probability, $\phi(p)$, for the combined B_1 and B_2 groups. The diagonal line represents the points where p and $\phi(p)$ are equal, and consequently, deviations from the diagonal line represent the magnitude of the difference between p and $\phi(p)$. Note that there is very little difference between the three groups for p values greater than .50 while for p values less than .50, there seems to be a definite order of the three groups with respect to the magnitude of the deviations, i.e., from A_1, A_2, to A_3, respectively. Also note that for all groups the deviations are much smaller for p values in the neighborhood of .50 than for p values away from .50.

These results seem to indicate that *Ss* are more accurate in estimating the probabilities of events under the following conditions: (*a*) when there are few possible outcomes or events *and* the probabilities are less than .50, and (*b*) when the probabilities are in the neighborhood of .50. These two generalizations are not inconsistent with the significant main effects and interaction for Variables A and P as found in the analysis of variance.

With regard to Variable P, the results of this study are consistent with the results found by Preston and Baratta (1948) who found that low probabilities were overestimated and high probabilities underestimated. With respect to Variable B, the results of the study were contrary to expectations. It was felt that estimating odds using numbers would differ from estimating odds using monetary values, i.e., the utility of money would influence the estimates of odds. It might be argued, however, that asking *S* to imagine himself in a betting situation is quite different from placing him in a betting situation where

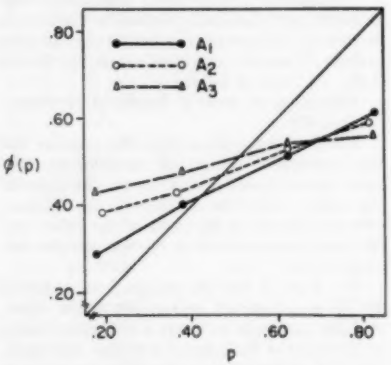


FIG. 1. Graph of subjective probability, $\phi(p)$, as a function of objective probability, p, for the A_1 (10 coins), A_2 (100 coins), and A_3 (500 coins) groups. Each point represents the mean subjective probability of 24 *Ss*.

he would actually play the bets, and thus, could win or lose his own money.

Summary

The primary purpose of this study was to determine factors which might influence an individual's estimate of the probability of events. Three variables were studied: (*a*) the number of outcomes or events which could occur in the game, (*b*) estimation using numbers vs. using monetary values, and (*c*) probability value. An analysis of variance design was used with 72 male undergraduate students as *Ss*.

The results indicated that *Ss* are least accurate in estimating the probability of events when the number of events or outcomes which can occur is large, and when probability values deviate from .50. This conclusion must be qualified, however, since the interaction between these factors was also found to be significant. With respect to the comparison between using numbers and using monetary values, under the conditions of this study, no significant differences were found.

References

Cohen, J., Dearnaley, E. J., & Hansel, C. E. Measures of subjective probability —Estimates of success in performance in relation to size of task. *Brit. J. Psychol.*, 1957, **48**, 271–275.

Davidson, D., Suppes, P., & Siegel, S. *Decision making; an experimental approach.* Stanford: Stanford Univer. Press, 1957.

Edwards, W. Probability preference in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

Preston, M. G., & Baratta, P. An experimental study of the auction-value of an uncertain outcome. *Amer. J. Psychol.*, 1948, **61**, 183–193.

(Received December 20, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

ACQUISITION OF PROBABILISTIC PAIRED ASSOCIATES AS A FUNCTION OF S-R_1, S-R_2 PROBABILITY [1]

JAMES F. VOSS,[2] CHARLES P. THOMPSON,[3] AND JAY H. KEEGAN [4]

Wisconsin State College, Eau Claire

This paper reports three experiments designed to explore probabilistic verbal learning. Paired-associate learning has typically involved S-R stimuli with one stimulus and one response word. Retroactive inhibition experiments have employed a stimulus with one response in original learning and a second response in interpolated learning (Bugelski, 1942). However, investigation of acquisition involving simultaneously competing responses has not been systematically investigated and is fundamental to a better understanding of complex learning and especially of verbal behavior, which is characteristically probabilistic (Osgood, 1954).

Probability learning studies have of course involved competing responses (Grant, Hake, & Hornseth, 1951). The present experiments extend the findings of probability learning in two ways: (*a*) verbal stimuli were employed; (*b*) six associations were concomitantly learned rather than one.

EXPERIMENT I

Method

A probabilistically modified paired-associate procedure was employed which involved association of two response words with one stimulus word for five of six stimulus words. The trials were constructed so that one response associated with a given stimulus occurred for a given percentage of trials of each 10-trial block while the second response occurred on the remaining trials of each 10-trial block. The probabilities of R_1 and R_2 for each S-R_1, S-R_2 association were varied as indicated in Table 1.

The stimulus words and their respective scale values are also presented in Table 1. The R_1 and R_2 words for each S word were permuted for one-half of the *Ss* (with the exception of the 1.00–.00 condition).

The stimuli were presented on strip film at a rate of 2 sec. per frame. A standard paired-associate procedure was employed (McGeoch & Irion, 1952) with one blank frame after each trial. Each film strip contained two 10-trial blocks and was presented for 120 trials. One-half of the *Ss* were started at each of the two 10-trial blocks. Starting point was orthogonal to response word permutation.

The probability conditions associated with each S-R_1, S-R_2 set of verbal material and the intertrial and intratrial sequences were assigned randomly. Fifteen *Ss*, students from psychology classes at Wisconsin State College, were randomly assigned to each condition ($N = 60$).

Instructions read to *S* were as follows: "A stimulus word will be presented to you on the wall in front of you. This presentation will be followed by the same word together with a response word. A series of such stimulus and response words will occur. Your task is to say the response word which you think will be presented with the stimulus word on all subsequent presentations of the same stimulus words. An example would be: the word *dog* will occur, then the words *dog-cat*. On each of the remaining presentations of the word *dog*, you would try to give the word you think will be associated with the word *dog*. Please note: (1) artificial words are used, so pronounce them the best you can; (2) say the response word that you think will occur with the stimulus word, even though you are not sure of it; (3) although you probably will not be able to get all the words correct, it is important that you respond as often as possible and try to get

[1] This investigation was supported by a research grant (M-1476) from the National Institute of Mental Health, United States Public Health Service. The authors also wish to thank Norman Anderson for his helpful comments.

[2] Now at the College of Wooster.

[3] Now at the University of Wisconsin.

[4] Now at the State University of Iowa.

TABLE 1

MEANS AND *SD*s OF SCALED NONSENSE WORDS
(1 = HIGH SIMILARITY TO ENGLISH; 7 = LOW SIMILARITY TO ENGLISH)

Probability Condition	Stimulus			Response		
	Word	Mean	*SD*	Word	Mean	*SD*
1.00–.00	YAWM	2.85	1.85	MINS	3.13	1.41
				KOSH	3.05	1.66*
.90–.10	DOCT	2.80	1.54	HUFT	3.30	1.71
				GAWS	3.23	1.80
.80–.20	BARP	2.85	1.87	MURT	3.05	1.50
				HOLS	3.10	1.85
.70–.30	PERS	2.83	1.40	JAWL	3.03	1.64
				GOSP	3.00	1.79
.60–.40	LIFS	2.88	1.70	KARN	3.05	1.61
				MENK	3.05	1.58
.50–.50	HALK	2.70	1.54	VORD	3.30	1.81
				BINT	3.30	1.71
IC-4 Pretraining words, Exp. 2	JORT	3.85	1.38	RAWZ	4.08	1.76
	BURZ	3.78	1.70	NIRK	4.30	1.87
				HARX	4.20	1.76
	HERF	4.03	1.34	VAMS	4.13	1.48
				FOLP	4.20	2.12
	SEMP	3.73	1.67	HIDZ	4.05	2.05
				ZAPS	4.05	1.67
	ROSP	4.13	1.76	SIRN	4.38	1.70
				TENX	4.38	1.82
	TARB	3.75	1.79	FESK	4.03	1.84
				BONT	4.05	1.79

* Used only in Exp. 3.

as many words correct as you are able to. Do you have any questions?"

Only one response by *S* was allowed for each stimulus word presentation. The response classes recorded for each stimulus word were: R_1, R_2, no response, and extraneous response. Hereafter, T_i responses refer to responses given by *S* of the particular response word which occurred with a probability i following the respective stimulus word.

Results

Response frequency.—Figure 1 presents response percentage as a function of 10-trial blocks with T_i response as parameter. The raw data of the .50 to .90 and the .10 to .50 conditions were analyzed independently due to the lack of independence of the R_1 and R_2 words. One .50 condition was arbitrarily assigned to each analysis. The 1.00 probability condition was omitted because of the relative lack of variance. The summary of the .50 to .90 analysis is presented in Table 2. (Significant

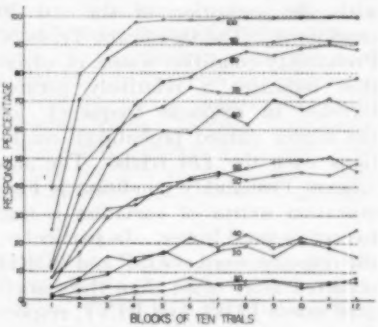


FIG. 1. Response percentage as a function of 10-trial block with T_i response as parameter.

TABLE 2

ANALYSIS OF VARIANCE OF THE .50–.90 PROBABILITY CONDITIONS: EXP. 1

Source of Variation	*df*	*MS*	*F*
Between *Ss*			
Permutation (P)	1	13	<1
Starting point (SP)	1	7	<1
P × SP	1	23	<1
Residual between *Ss*	56	59	
Within *Ss*			
Trials (T)	11	1054.4	190.98*
Prob. condition (PC)	4	2206.5	401.18*
T × PC	44	21.9	3.98*
T × P	11	1.1	<1
T × SP	11	2.2	<1
PC × P	4	636.0	115.64*
PC × SP	4	76.8	13.96*
T × P × SP	11	4.7	<1
PC × P × SP	4	25.4	4.62*
T × PC × P	44	5.9	1.07
T × PC × SP	44	5.0	<1
T × PC × P × SP	44	2.4	<1
Residual within *Ss*	3304	5.5	

* $P = .01$.

sources of variation are the same in the .10 to .50 analysis with the exception that the T × PC × P interaction is significant.)

Duncan (1955) range tests performed on the means of the significant probability conditions indicated that all probability conditions significantly differed from each other (.05 level) with the exception of the .10–.20 conditions. The significant Trials × Probability condition source of variation indicates a relatively greater increase in response frequency for the higher valued probability conditions over the 120 trials. The significant Probability condition × Permutation source of variation is due to certain word biases. In particular, the response words GOSP and VORD occurred more often than the associated words JAWL and BINT, respectively. The means of the Probability condition × Starting point interaction yielded no systematic trends.

Asymptotic performance.—Analyses of variance performed on the .50 to .90 and .10 to .50 conditions for Trial Blocks 11 and 12 indicated that no source of variation containing trials is significant. An asymptote was therefore assumed.

Figure 2 presents Block 12 T_i response percentage as a function of S-R probability. The data indicate that response percentage was above the "matching" line for the .60–.90 S-R probabilities and below for the .10–.40 S-R probabilities. Modified analyses of variance (Grant, 1955) employing orthogonal polynomials indicate significant linear and quadratic components (.50 to .90: linear, $F = 136.41$; quadratic, $F = 12.55$; .10 to .50: linear, $F = 111.67$, quadratic, $F = 14.82$; 1 and 295 *df* for all *F* ratios). The quadratic component indicates that response frequency deviates significantly from the stimulus probability. It should be noted that all permutation and starting point response frequencies for each probability condition occurred on the respective side of the "matching" line.

Rate of acquisition.—Since acquisition curves and asymptotic performance do not yield measures of rate of learning where rate is defined as the proportion of acquisition from the zero level to the asymptotic level that occurs on successive trials, a procedure was employed analogous to the estimation of θ used by Estes and Straughan (1954). The modification (Voss, 1958) involved estimation of the asymptotic performance level from the data since no a priori assumptions were made regarding asymptotic performance. Moreover, θ' is employed to designate rate to avoid the theoretical implications of θ.

Table 3 presents learning rate

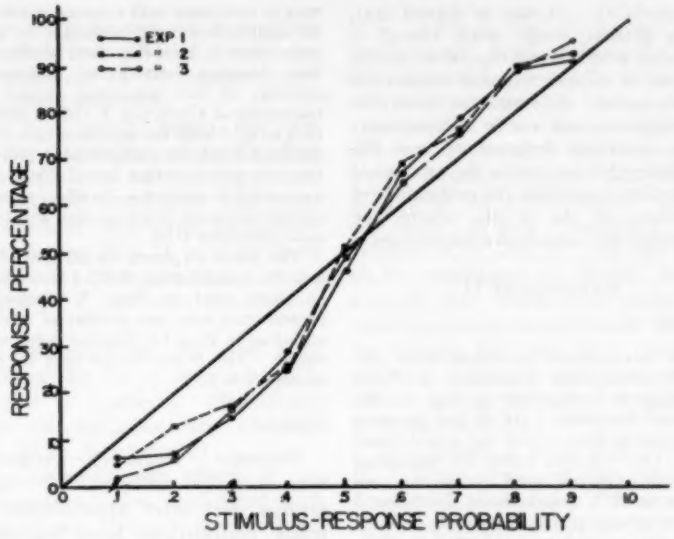


FIG. 2. Response percentage as a function of S-R probability for the Trial Block 12 data of Exp. 1, 2, and 3.

values for the probability groups with asymptotic estimates based upon the last 20 trials. The values indicate that rate generally increases as a function of probability condition although major inversions occur at the .70 and .30 conditions. No run or alternation tendencies were uniquely present in the .70–.30 condition to account for the inversion. The inversion was instead likely due to the confounded verbal material.

TABLE 3

LEARNING RATE, θ', FOR THE SAME VERBAL MATERIAL AT THE RESPECTIVE PROBABILITY CONDITIONS IN EXP. I AND II

Probability Condition	θ'	
	Exp. 1	Exp. 3
.90	.099	.077
.80	.040	.061
.70	.050	.045
.60	.042	.031
$.50_1$	.039	.037
$.50_2$	.041	.039
.40	.030	.029
.30	.064	.028
.20	.033	—
.10	.023	.026

Discussion

The verbal aspects of the present experiment initiate certain problems not found in other probability learning situations. In particular, response word biases occurred in the present experiment despite the use of individually scaled nonsense words. Intralist interference would likely produce word bias, even though the use of scaled S-R_1, S-R_2 pairs of words would probably minimize bias.

Acquisition is likely influenced by the interference between respective R_1 and R_2 words and the other response words. Gardner (1957), in a situation involving one stimulus and three responses, found that the most frequently associated response yielded response percentages considerably above the respective stimu-

lus probability. It may be argued that, in the present study, even though a somewhat arbitrary asymptote occurred, each pair of correct response words were not adequately differentiated from the remaining response words. Experiment II was therefore designed to test the hypothesis that increasing the amount of information concerning the probability of occurrence of the words affects the acquisition and asymptotic performance.

Experiment II

Method

Four instructional conditions were employed. Instructional Condition 1 (IC-1) had the same instructions as Exp. I. Instructional Condition 2 (IC-2) had the same instructions as Exp. I with the added statement, "Two response words are associated with each stimulus word except for one stimulus word." Instructional Condition 3 (IC-3) employed pretraining which involved presentation on each trial of both S-R_1, S-R_2 associations used in the experimental test situation. Instructions to *S* in pretraining were to anticipate both response words. The six associations were presented in random order over a recurring 20-trial film. The test situation immediately followed the criterion of two successive correct trials. Instructional Condition 4 (IC-4) was identical to IC-3 with the exception that different nonsense words were employed in pretraining than the test situation (see Table 1). IC-4 was added to determine the effect of previous nontest nonsense word learning and to serve as a control for IC-3.

The words employed for IC-1, IC-2, IC-3, and the test situation of IC-4 were the same as those used in Exp. I. Permutation, presentation rate, and number of trials were varied as in Exp. I. Starting point was not varied. The *N* was 12 for each IC permutation ($N = 96$).

Results

Response frequency.—Acquisition was initially different for the IC groups, but after approximately 40 trials, performance level became virtually equal for the IC groups. Figure 3 therefore presents the per-

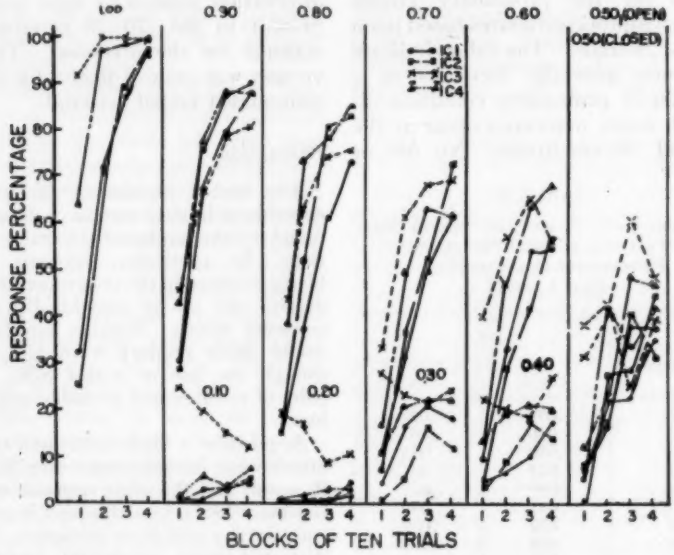


Fig. 3. Response percentage as a function of the first four 10-trial blocks with T_i responses and instructional condition as parameters.

centage T_i responses as a function of the first four 10-trial blocks for the parameters of T_i responses and instructional condition.

The IC-3 group began the test series by responding relatively close to the 50% level for both responses. Warm-up and familiarity with nonsense words in pretraining apparently enabled Group IC-4 to produce more early trial T_i responses than Groups IC-1 and IC-2, but fewer than Group IC-3.

The summary of the raw data analysis of the .50 to .90 condition is presented in Table 4. (Significant sources of variation are identical in the .10 to .50 analysis.)

The nonsignificant effect of instructional condition considered with the significant Trials × Instructional condition interaction indicates that T_i response frequency was greater for Groups IC-3 and IC-4 during early trials, but Groups IC-1 and IC-2 reached the IC-3 and IC-4 performance level sufficiently early in trials so that no over-all instructional difference occurred.

TABLE 4

ANALYSIS OF VARIANCE OF THE .50–.90 PROBABILITY CONDITIONS: EXP. II

Source of Variation	*df*	*MS*	*F*
Between *Ss*			
Permutation (P)	1	918.0	19.21*
Instructional cond. (IC)	3	79.0	1.65
P × IC	3	243.7	5.10*
Residual between *Ss*	88	47.8	
Within *Ss*			
Trials (T)	11	1154.63	232.79*
Prob. condition (PC)	4	2819.25	568.40*
T × PC	44	12.41	2.50*
T × P	11	31.27	6.30*
T × IC	33	27.24	5.49*
PC × P	4	515.50	103.93*
PC × IC	12	37.58	7.58*
T × P × IC	33	3.42	<1
PC × P × IC	12	80.42	16.21*
T × PC × IC	132	3.95	<1
T × PC × P	44	5.09	1.03
T × PC × P × IC	132	2.24	<1
Residual within *Ss*	5193	4.96	

* $P = .01$.

Inspection of the Instructional condition × Probability condition interaction means indicates that Groups IC-3 and IC-4 responded with a relatively greater frequency than Groups IC-1 and IC-2 at the probability conditions of .50–.50, .60–.40, and .70–.30. These three probability conditions were most difficult to learn for Groups IC-1 and IC-2, and pretraining yielded a relatively greater advantage for Groups IC-3 and IC-4 at these conditions.

Elimination of word bias in Group IC-3 likely produced the significant Instructional condition × Permutation and Instructional condition × Permutation × Probability condition interactions. Moreover, even on Trial Block 12 associations of both permutations have virtually equal frequency in IC-3. Groups IC-1, IC-2, and IC-4 demonstrated the word biases previously found. The IC-3 and IC-4 findings imply that the training with particular words is being transferred in IC-3 and not the "set" for relationships of probabilistic responses.

Asymptotic performance.—Analyses of variance (.10 to .50 and .50 to .90) of Trial Blocks 11 and 12 yielded no significant source of variation containing trials. An asymptote was therefore assumed.

Figure 2 presents T_i response frequency as a function of S-R_1, S-R_2 stimulus probability for Block 12. Analyses of the Block 12 data indicated that the linear and quadratic components of the significant probability condition source of variation are significant in both the .50 to .90 and .10 to .50 analyses (.50 to .90: linear, $F = 195.81$; quadratic, $F = 11.78$; .10 to .50: linear, $F = 201.92$; quad-

ratic, $F = 7.06$; 1 and 475 *df*). All instructional condition means for the .60 to .90 and .10 to .40 conditions occur on the respective side of the "matching" line.

Discussion

Pretraining in IC-3 had three effects: (*a*) Results of Exp. 1 and 2 indicated that asymptotic performance is not a function of the degree of learning. (*b*) A relatively transient effect of pretraining was found which indicated test performance is at first a function of pretraining conditions, but rapidly becomes a function of the test probabilities. (*c*) Word bias was eliminated.

The findings of the remaining groups indicate that pretraining on nontest nonsense words has only a transient effect in early learning and knowledge of the number of responses has no effect upon performance.

Experiment III

Method

Experiment III was designed to study concomitant acquisition of six probabilistic associations of the same probability value. Moreover, use of the same verbal material allowed for comparison between the S-R_1, S-R_2 association employed at the same probability values of Exp. I and III.

Five probability conditions were employed with the same values as those of Exp. I and II with the omission of the 1.00–.00 condition. The S-R_1 and S-R_2 probabilities were the same for all verbal associations within the given probability group. Other conditions of Exp. III were comparable to Exp. I and II.

Ten *Ss* were assigned randomly to each probability condition permutation, or $N = 100$.

Results

Response frequency.—Acquisition curves for the five probability conditions are similar to those presented in Fig. 1 and may be found elsewhere (Voss, 1958). A summary of the .50 to .90 analysis is presented in Table 5. The only additional significant source of variation found in the .10 to .50 analysis is Permutation. The significant Probability condition, Trials, and Trials × Probability condition interaction sources of variation demonstrate generalization of the results found in Exp. I and II. Duncan range tests indicated that all probability conditions differed significantly from each other.

TABLE 5

Analysis of Variance of the .50–.90 Probability Conditions: Exp. III

Source of Variation	*df*	*MS*	*F*
Between *Ss*			
Prob. condition (PC)	4	4219.5	33.38*
Permutation (P)	1	2.0	<1
PC × P	4	135.3	1.07
Residual between *Ss*	90	126.4	
Within *Ss*			
Trials (T)	11	2338.4	441.21*
Verbal material (VM)	5	86.8	16.38*
T × VM	55	8.0	1.51
T × PC	44	36.1	6.81*
T × P	11	24.6	4.64*
VM × PC	20	46.5	8.77*
VM × P	5	245.0	46.23*
T × PC × P	44	4.0	<1
T × VM × PC	220	5.1	<1
T × VM × P	55	4.4	<1
VM × PC × P	20	41.4	7.81*
T × VM × PC × P	220	4.0	<1
Residual within *Ss*	6390	5.3	

* $P = .01$.

Since verbal material was not confounded with probability condition in the present experiment, variability of verbal material may be analyzed in greater detail. Difficulty increased in the following stimulus word order: .50 to .90, DOCT, YAWM, BARP, HALK, PERS, and LIFS; .10 to .50, YAWM, DOCT, HALK, PERS, LIFS, and BARP. The nonsignificant Trial × Verbal condition interaction indicates that the relative difficulty of verbal material was virtually fixed over trials. Inspection of the means indicated that

no particular verbal stimuli were systematically more difficult as a function of S-R_1, S-R_2 probability.

Asymptotic performance.—Analyses of variance of the .10 to .50 and .50 to .90 conditions in Trial Block 11 and 12 revealed that no significant source of variation containing trials. An asymptote was therefore assumed.

Figure 2 presents the Block 12 data. Probability condition is a significant source of variation and subsequent analyses revealed that the linear component is significant in both analyses (.50–.90: $F = 214.75$; .10–.50: $F = 161.56$; 1 and 90 *df*). The quadratic component is significant in only the .10–.50 analysis (.50–.90: $F = 2.63$; .10–.50: $F = 7.77$; 1 and 90 *df*). The results of the .50 to .90

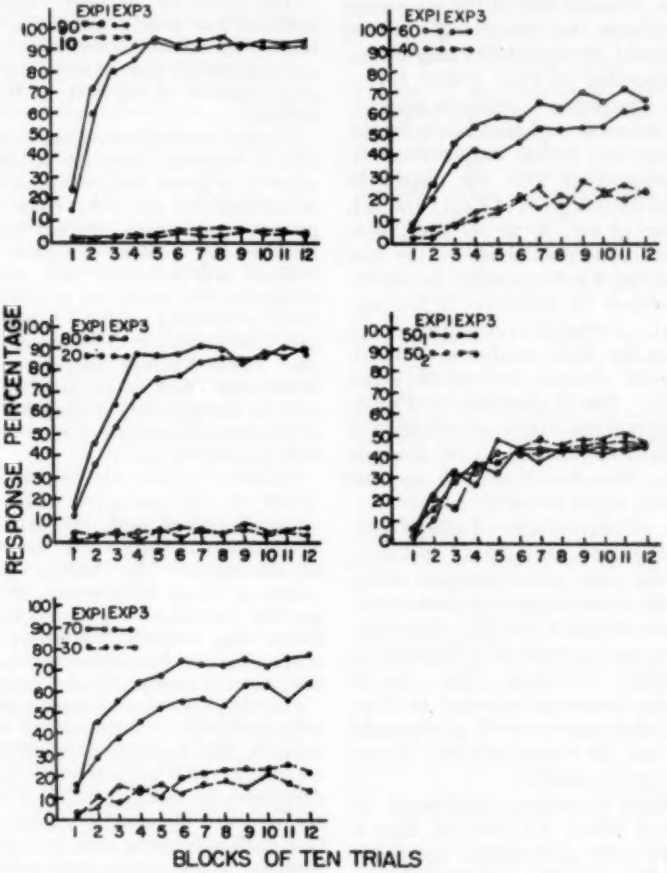


FIG. 4. Response percentage as a function of 10-trial blocks with T_i response as parameter for the one identical verbal material condition associated with the respective probability condition of Exp. I and III.

analysis do not necessarily imply that response frequency occurred at the "matching" rate. The mean frequencies of response for the .60–.90 groups instead occurred above the "matching" line.

Verbal material results however complicate the asymptotic situation. The six verbal material Trial Block 12 means occur above the "matching" line for the .90 and .80 stimulus probabilities, whereas four of the six means occur above the "matching" line in the .70 and .60 probability conditions.

Comparison of Exp. I and III.—Figure 4 presents T_i response percentage as a function of trial block for the one identical verbal material condition association with the respective probability condition of Exp. I and III. Analyses of the .50 to .90 and .10 to .50 conditions performed on the raw data of Fig. 4 indicate that the experiment source of variation is not significant. Probability condition is significant for both analyses (.50–.90: $F = 65.93$; .10–.50: $F = 49.24$; 4 and 390 *df*). The Experiment × Probability condition source of variation is significant for the .50 to .90 analysis (.50 to .90: $F = 3.49$; .10 to .50: $F = <1$; 4 and 390 *df*).

Rate of acquisition.—Table 3 presents the rate of learning measure, θ', for the same verbal material which occurred at the respective probability condition of Exp. I and III. The Exp. III θ' values increase as a function of probability condition. No .70–.30 condition inversion is found in Exp. III. A decrease of the $T_{.20}$ responses in the last 20 trials provided for no meaningful measure.

Analyses of variance performed on the Trial Block 12 data of Fig. 4 revealed only probability condition as significant. This result indicates that although the interaction of Probability condition and Experiment is significant in the .50 to .90 analysis, and although some differences occur in the θ' values presented in Table 3, asymptotic performance was not significantly different between the groups having the same verbal material employed at a given probability condition in Exp. I and III.

Discussion

The results of the three experiments indicate that over a number of conditions, response frequency, learning rate, and asymptotic performance level varied as a function of the $S\text{-}R_1$, $S\text{-}R_2$ probabilities.

Further research on response competition is required where the R_2 response word in a given $S\text{-}R_1$ $S\text{-}R_2$ association is omitted on the $S\text{-}R_2$ trials of the present experiments. Partial reinforcement studies (Jenkins & Stanley, 1950) indicate that relatively little difference in acquisition occurs as a function of $S\text{-}R_1$ probability values between .50 and 1.00 where the $S\text{-}R_2$ probability is .00. These findings suggest that the differential effect of probability condition in the present experiments is due to response competition rather than the $S\text{-}R_1$ probability per se.

Intralist response interference, as revealed by intrusions of intralist responses, occurred early in learning and diminished as performance approached an asymptote. This finding and the results of Group IC-3 in Exp. II, where intralist intrusions were minimized, indicate that intralist interference influences response frequency and learning rate, but not asymptotic performance.

Learning rate was a function of $S\text{-}R_1$, $S\text{-}R_2$ probability. Inspection of Table 3 suggests that extreme probabilities are required in order to yield relatively large differences in learning rate. Two hypotheses are therefore suggested. The first is that learning rate is a sigmoid function of $S\text{-}R_1$, $S\text{-}R_2$ probabilities in similar competing response situations. The second hypothesis is that acquisition is a function of the degree of discrimina-

tion between the probabilistic associations of the list. This hypothesis assumes that discrimination of S-R_1, S-R_2 events increases as the probabilities vary from .50–.50 to .90–.10. Little evidence however is available on probabilistic psychophysical judgments (Attneave, 1953; Clarke, 1957).

Further research is required in order to determine the effect of the interaction of number of associations and probability of associations upon acquisition. However, since rate of learning is more rapid for the S-R_1 associations approaching the 1.00 probability, it is expected that acquisition is a function of the proportion of associations learned with the greater S-R_1 probabilities. In addition, the comparison of Exp. I and III indicates that at least for the .50 to .90 conditions, acquisition is a function of the probability of the concomitantly learned associations.

Asymptotic performance level was not related to degree of learning or the probability of concomitantly learned associations. Further experimentation is required in order to determine when asymptotic response level varies.

The results of the present experiments are limited in that verbal material was confounded with probability condition in Exp. I and II. However, the results of Exp. III indicate that although significant differences occurred between verbal material, the verbal material variable involved a relatively small error in the response frequency-stimulus probability relationship.

Summary

Three experiments were performed on the acquisition of S-R_1, S-R_2 probabilistic associations. Six probabilistically modified paired associates were presented for 120 trials. The six probabilities were: 1.00–.00; .90–.10; .80–.20; .70–.30; .60–.40; .50–.50. The variable of stimulus information was studied in Exp. II. Stimulus probability was made a between-*S* variable in Exp. III.

The results indicate that response frequency is an increasing function of stimulus probability. Learning rate increased as a function of S-R probability, although large differences occurred only at extreme probability values. Asymptotic performance level was virtually fixed over the three experiments above the respective S-R probabilities for .50 to .90 conditions and below the complementary S-R probabilities for the .10 to .50 conditions.

The results were related to S-R_1, S-R_2 response competition and intralist interference. Hypotheses suggested included that learning rate is a sigmoid function of S-R probability, and is also related to the discrimination of S-R_1, S-R_2 probability differences.

References

Attneave, F. Psychological probability as a function of experienced frequency. *J. exp. Psychol.*, 1953, **46**, 81–86.

Bugelski, B. R. Interference with recall of original responses after learning new responses to old stimuli. *J. exp. Psychol.*, 1942, **30**, 368–279.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29**, 715–720.

Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, **2**, 171–189.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Gardner, E. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

Grant, D. A. Tests subsequent to over-all analysis of variance and the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

Osgood, C. E. (Ed.). Psycholinguistics. *J. abnorm. soc. Psychol.*, 1954, **49**, (Suppl., Part 2).

Voss, J. F. Language behavior and stimulus probability. Rep. M-1476, Nat. Inst. of Ment. Health, Bethesda, Md.: 1958.

(Received December 22, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

SECONDARY AND GENERALIZED REINFORCEMENT IN HUMAN LEARNING[1]

FREDERICK H. KANFER[2] AND JOSEPH D. MATARAZZO

Purdue University *University of Oregon Medical School*

A considerable burden has been placed by theorists on the concept of generalized reinforcement in attempts to explain the power of socially mediated reinforcers, such as "good," attention, or approval. Skinner (1953) describes the generalization of reinforcing properties of a conditioned reinforcing stimulus as a consequence of pairing the stimulus with numerous primary reinforcers. He suggests that generalized reinforcement is more effective because it is less dependent on the momentary deprivation of the organism and the same response may precede a variety of reinforcements on different occasions and under different deprivations, thus increasing the probability of the future occurrence of the response. Furthermore, the wide variety of rewards which have followed a gener. lized reinforcer decreases the likelihood that satiation has taken place after several instances of the sequence. Utilization of tokens as generalized reinforcers with distinct physical characteristics, should further enhance the effect, since "these permit a sharper contingency between behavior and consequence" (Skinner, 1953, p. 79).

The purpose of the present study is to investigate the relative effectiveness of a *discrete* secondary reinforcer vs. a *generalized* secondary reinforcer; i.e., one which has been paired with only one consumable reward vs. one which has been paired with different consumable rewards.

[1] This research was supported in part by Research Grants M-2027 and M-1938 from the National Institute of Mental Health, United States Public Health Service.

[2] The data were collected while the first author was visiting professor in the University of Oregon Medical School, Summer 1958.

METHOD

Subjects.—The *Ss* were 54 female student nurses who volunteered for participation in this study. All were naive about psychological experimentation and the purpose of the study.

Apparatus and procedure.—The *S* sat facing a memory drum. To her right was a 12-in. high wooden partition mounted on a table which allowed *E* to observe *S*, but hid *E*'s recording of data. Inserted within the partition was a tin chute which extended from *E*'s side of the partition to *S*'s side, where it protruded 4 in. from the partition. A token was dropped into the tin chute by *E* from his side and emerged from the outlet on *S*'s side into an open box in full view of *S* whenever she gave a correct response. On a table between *E* and *S*, and in full view of *S* at all times, was a tray containing a quantity of the reward(s) for which she could exchange the tokens. Blue poker chips served as tokens. The rewards consisted of candy bars, sample packs of cigarettes, and sample flasks of hand lotion. The rewards were approximately equal in commercial value (about 5 cents each) and were selected after approximately equal preference was expressed for these rewards among a sample of female nurses.

Two lists of paired associates were used as learning tasks. The first list consisted of six pairs of adjectives which had average similarity ratings of .15 or less, as reported by Hilgard (1951, p. 552) after a study by Melton and Saffir. The second list contained six pairs of nonsense syllables taken from Hilgard's list of syllables with low association values (1951, p. 545). Stimulus presentations and intertrial intervals of 2 sec. were used on both tasks. A trial is defined as the presentation of one pair.

The *Ss* were randomly assigned to three groups which differed only with respect to the eventual reward for which poker chips could be exchanged. All *Ss* were given the

usual instructions for paired-associate learning, and were told that they would be given one poker chip for each correct anticipation.

This ended the instructions for the Control Group (C). The *Ss* in the secondary reinforcement group (SR) were divided into three equal subgroups. The subgroups differed only in that the first obtained *only* cigarettes as a reward, the second obtained *only* candy bars, and the third obtained *only* hand lotion. For the entire SR group, the following was added: "At certain intervals during the experiment you may exchange the chips for bottles of hand lotion (packs of cigarettes, candy bars). Two chips may be traded for one bottle of hand lotion (one pack of cigarettes, one candy bar). You may keep the items which you have earned."

The generalized reinforcement group (GR) was told: "At certain intervals during the experiment you may exchange the chips for candy bars, packs of cigarettes, or bottles of hand lotion. Two chips may be traded for any one of these. You will have your *choice among the three* each time and you may keep what you choose.

Thus, while *Ss* in Group C were not able to exchange their poker chips for any reward, and *Ss* in Group SR could exchange theirs for only one type of reward, *Ss* in Group GR could exchange their chips for some of each of the three rewards.

After each *S* had seen the respective rewards which she could expect and had indicated that she understood the instructions, learning of the adjective pairs was begun. All *Ss* received one poker chip for every correct association. The *Ss* continued until 20 correct associations were given. By this method each *S* obtained 20 poker chips during this part of the experiment. After reaching the criterion on the first learning task, *S* was allowed to exchange his tokens for whatever rewards were appropriate to his group. In Group C the tokens were collected and no further comments were made. In the reward groups, *Ss* were encouraged to handle rewards themselves and were asked to keep the rewards in a designated place next to the memory drum in front of them. The *E* then inserted the nonsense syllable list in the memory drum and proceeded to give the instructions for the second task. For all *Ss* an interval of 4 min. was allowed between the end of the first learning task and the presentation of the second (paired nonsense syllables) learning task. All *Ss* were given the following instructions during the interval between learning Tasks 1 and 2: "For the next part, paired nonsense syllables will be used instead of adjectives (sample trial). Notice that the second syllable must be spelled, not pronounced. In all other ways your task will be the same as before. For every correct association, you will be given one poker chip. The chips will be turned in again at the end of the learning period.

The *Ss* were encouraged to believe that these tokens would again be redeemable by leaving full trays of rewards in front of them throughout the entire experiment. No questions were answered concerning redemption of the chips. The nonsense list was presented until *S* had given 20 correct anticipations. For all groups, one poker chip was given for each correct response. The poker chips were collected after the completion of the second learned task and *Ss* in Groups GR and SR were questioned concerning their use of the rewards which they had obtained. For instance, an *S* in Group SR who had obtained cigarettes would be questioned about her smoking habits, whereas an *S* who had been given free choice of all three rewards and had selected candy, cigarettes, and hand lotion would be asked about her use of each product and her relative preference for each of the rewards.

Results

Table 1 summarizes the mean number of trials (presentations of S-R pairs) on both learning tasks and the mean difference between the two tasks. The difference scores were obtained by subtracting each *S*'s score on the second task from his score on the first task.

The analysis of the data is presented in Table 2. The experimental groups, despite differences in instructions, did not differ in learning the adjective pairs (Task 1). It should be noted that this task was carried out before any actual exchange of tokens for the rewards which Groups SR and GR had been promised. On the nonsense syllable learning (Task 2), the groups differed significantly from each other ($F = 4.33$). Group C required the largest number of trials to the criterion, while *Ss* in Group GR showed fastest learning. As expected, the two learning tasks

TABLE 1

MEAN NUMBER OF TRIALS ON BOTH TASKS AND MEAN DIFFERENCE SCORES BETWEEN TASKS

Groups $N = 18$	Task 1		Task 2		Differences	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Control	47.06	12.94	70.89	20.83	23.83	19.58
SR	45.11	15.86	58.39	15.24	13.28	21.28
GR	46.89	14.79	52.67	17.64	5.78	17.63

differed significantly from each other ($F = 26.40$). The significant Task by Group interaction ($F = 3.54$) indicates that the treatment resulted in differential performance changes among the groups on the two tasks. A breakdown of the effects of the task on the three groups is also presented in Table 2. Group GR shows no significant change in its mean performance on the two tasks. The other groups, however, require significantly more trials on the nonsense syllable list than on the adjective list. The data indicate, therefore, that only generalized reinforcement was sufficiently effective to facilitate learning on Task 2 to the point that mean number of trials to criterion was no higher than for the relatively easier adjective list.

In addition to the over-all F tests, the extent of the effectiveness of generalized reinforcement over secondary reinforcement had been of interest in this study. Therefore, this specific hypothesis was tested by an F test on the nonsense syllable scores of Groups SR and GR. The resulting F of 1.03 (1 and 34 *df*) indicated that these two groups did not differ significantly from each other.

Since group SR had three subgroups which received different rewards, an analysis of their performance was carried out (Table 3). An F of 1.18 indicated that the type of reward did not significantly affect learning. Furthermore, a nonsignificant F for Tasks by Subgroups was obtained. Thus, the type of reward did not show any interaction with the tasks.

TABLE 2

ANALYSIS OF VARIANCE ON ALL SCORES

Source	*df*	*MS*	*F*
Between *Ss*	53		
Groups (G)	2	843.53	2.33
G in Task 1	2	20.91	<1
G in Task 2	2	1563.13	4.33*
Pooled *Ss* in G	51	361.35	
Within *Ss*	54		
Tasks (T)	1	5518.37	26.40**
T in Gp. C	1	5112.25	24.45**
T in Gp. SR	1	1586.69	7.59**
T in Gp. GR	1	300.44	1.44
T × G	2	740.51	3.54*
T × *Ss* in G	51	209.05	
Total	107		

* $P < .05$.
** $P < .01$.

TABLE 3

ANALYSIS OF VARIANCE ON SUBGROUPS IN GROUP SR

Source	*df*	*MS*	*F*
Between *Ss*	17		
Subgroups (SG)	2	292.00	1.18
Pooled *Ss* in SG	15	247.55	
Within *Ss*	18		
Tasks (T)	1	1586.69	7.97*
T × SG	2	199.14	<1
T × *Ss* in SG	15	267.50	

* $P < .05$.

Discussion

Despite the apparent trends, the major hypothesis concerning the greater effectiveness of generalized reinforcement over secondary reinforcement has not been clearly substantiated. However, the experimental findings point to several issues which must be clarified in further experimentation before the implication of these results can be evaluated.

The present study utilized tokens as an intermediate step between the response and the reward. Although tokens have the advantage of reducing satiation effects, poker chips may themselves have some initial generalized incentive value and therefore act as generalized reinforcers for all groups.

The rewards were equated for their commercial value and also for preferences among an initial group of pilot *Ss*. Nevertheless, *Ss* in Group GR showed some preferences for one reward (hand lotion) over the other two. Although the type of reward in itself did not have a major effect, as indicated by the test on the three subgroups in Group SR, a preference for one type of reward over the others would result in lessened effectiveness of a generalized reinforcer. In further research in this area, it would seem advisable to control individual preferences for *Ss*, in order to sustain equal motivation in all *Ss* in a GR group or treat it as an additional independent variable.

Although Groups SR and GR did not differ significantly from each other on the second task, the data also indicate that only Group GR showed sufficient improvement in learning the nonsense syllable list, when compared with the relatively easier adjective list to yield performance scores on these two tasks which did not differ from each other. Thus, greater individual gains in Group GR may reflect a trend which is not sufficiently strong when tested only in a comparison of Groups SR and GR on Task 2. The statistical test on performance of Groups SR and GR in Task 2 only is, of course, less sensitive than the over-all analysis since the latter is based on an error term which is corrected for the individual consistencies of behavior on Tasks 1 and 2.

The instructions themselves did not affect group performances on Task 1. Both rewarded groups did significantly better only after they had been allowed to exchange their tokens for the rewards. It would appear then, that the actual pairing of the token with the promised reward was necessary to provide differential reinforcing properties to the tokens. As Skinner (1953) has indicated in his discussion of secondary and generalized reinforcement, special value of these reinforcing stimuli might be attributable to their potential for maintaining a high motivational level. If such motivational effects differed between Groups SR and GR, within-group variability with respect to the specific reward would become an important masking factor. Only if the rewards are indeed equally desirable would the potential for choosing several rewards lessen the effect of satiation and thereby maintain a more consistently high motivational level in Group GR.

Considering the direction of differences between the groups on Task 2, as well as the significantly greater improvement in Group GR from Task 1 to Task 2, failure to find full support for the experimental hypothesis by statistical tests, may indicate that further control over individual reward values may be necessary to provide sufficient opportunity for generalized reinforcers to become fully effective.

Summary

The study investigated the effectiveness of generalized reinforcement in learning. Fifty-four female *Ss* were assigned to three groups. All *Ss* first learned a paired adjective list and were given poker chips for correct responses. Twenty correct responses terminated learning. All *Ss* then learned a second task, a nonsense syllable list, to the same criterion of 20 correct responses. The generalized reinforcement group (Group GR) was instructed that tokens would be exchanged for one of three types of rewards.

Such exchange was allowed between Tasks 1 and 2. The secondary reinforcement group (SR) was similarly rewarded at the end of the first task, but randomly assigned to only one type of reward. The control group (Group C) could not exchange tokens for any reward. The results indicated that the groups did not differ from each other in learning the first task. On Task 2 Group C required more trials than Group SR, Group GR learned Task 2 in fewest trials. The over-all F test was significant ($P < .05$) but the difference between Groups SR and GR was not significant. Considering the effect of learning Task 1 on Task 2, it was found that performance scores differed significantly for Groups C and SR, whereas no difference in the number of trials to criterion was found for Group GR on Tasks 1 and 2. The results suggest that generalized reinforcement is somewhat more effective than secondary reinforcement in human verbal learning.

REFERENCES

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

(Received December 22, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

SCALES OF APPARENT FORCE [1]

JOSEPH C. STEVENS AND JOEL D. MACK

Harvard University

The purpose was to explore the functional relations between the subjective magnitude and the physical magnitude of mechanical forces experienced by an observer (*O*). There were four primary objectives: (*a*) to obtain a ratio scale for the apparent magnitude of a force exerted by *O* on a hand dynamometer, (*b*) to obtain a ratio scale for the apparent magnitude of a force applied by the experimenter (*E*) to the palm of *O*'s hand, (*c*) to compare and check these two scales of apparent force by having *O* exert force on the hand dynamometer to match the force applied to his palm, and (*d*) to compare, by a similar matching procedure, the subjective scale of exerted force and the "veg scale" for the subjective heaviness of lifted weights. In addition, a secondary objective was to determine the relation of the category scale of exerted force to the ratio scale of the same continuum.

What was perhaps the first attempt to scale the subjective magnitude of force was reported by Fullerton and Cattell (1892). They asked two *O*s to exert, on a hand dynamometer, forces twice as great or one-half as great as a standard force. They were puzzled by the fact that on the average the *O*s did not double or halve the physical magnitude of the standard, and they abandoned this pursuit. Recently, S. S. Stevens (1957) pointed out that the meager results of Fullerton and Cattell are roughly consistent with the veg scale, and he suggested that both the apparent magnitude of muscular pull and the apparent heaviness of lifted weights might grow as power functions of stimulus magnitude. The present studies undertake to repeat the experiment of Fullerton and Cattell and to extend the investigation along several lines.

[1] This research was supported partly by a grant from the National Science Foundation, partly by Contract Nonr-1866(15) with the Office of Naval Research (Project Nr142-201, Report PNR-225). Reproduction for any purpose of the U. S. Government is permitted.

FORCE OF HANDGRIP

Apparatus

In all the experiments *O* pulled on a handle with his right hand. For one set of experiments there was virtually no movement of the handle upon the application of force. The handle was connected, as shown in Fig. 1, to a calibrated Dillon Tensile-Type Force Gauge, graduated in 1-lb. units from 0 to 100 lb. A displacement of the handle through 0.039 in. caused a full 100-lb. deflection of the dial. The force gauge was connected with the handle and with the supporting table by means of ball-bearing joints, which tended to align the gauge in the direction of the force exerted on the handle.

For another set of experiments the same handle was connected by means of a steel wire and a fixed pulley to a suspended balance,

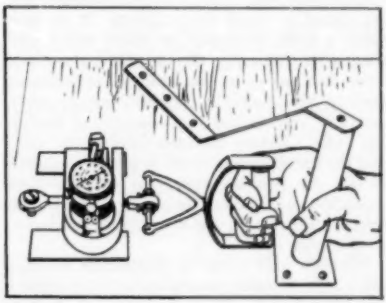


FIG. 1. Showing the parts and arrangement of the dynamometer.

whose dial was graduated in 1/10-lb. units from 0 to 40 lb. Calibration showed that the displacement of the handle was approximately proportional to the pull exerted on it, and that a 40-lb. pull required a handle displacement of approximately 1.5 in.

Procedure

In order to avoid bias that could possibly result from basing a scale on a single kind of judgment, three scaling methods were used: ratio production, magnitude production, and magnitude estimation. In addition, the method of category production was used to obtain a 7-point category scale. (For description of these methods, see S. S. Stevens, 1958.)

The possibility that apparent movement of the handle might influence the perception of force was investigated by using both the apparatus that required *O* to pull the handle over perceptible distances, and the apparatus that essentially eliminated movement. The ratio production and the category judgments were obtained with the movable handle only, but magnitude estimates and magnitude productions were obtained with both setups.

The *O*s were right-handed men, mostly students in the university. Some participated in a single session, others in more than one. Before a session, *O* read written instructions appropriate to the experiments, abbreviated versions of which appear below. Since one of the objectives was to obtain stable individual functions, a fairly large number of judgments was often required. Consequently *O* was encouraged to rest whenever he felt tired.

Ratio production.—The *O* was asked to exert successively two forces that appeared to be in 2 to 1 relation to each other. The instructions were: "In this experiment you are asked to exert on a handle at various times a light, moderate, or heavy force, and then after releasing the handle to exert a force that seems to you one-half or twice as great. "One-half" or "twice" means what seems to you to feel like one-half or double the force. Before the experiment, you may pull the handle a few times to get accustomed to it. As you proceed, you will be told in each case whether to exert a moderate, light, or heavy pull, and also whether to halve or to double it."

Five *O*s first made 15 judgments of one-half apparent force, rested for a few minutes, and then made 15 judgments of double apparent force. Another five *O*s performed the same judgments in reverse order. Each group of 15 judgments contained five judgments in each of three blocks specified by *E* as light, moderate, or heavy. The sequence of blocks was varied from one *O* to another.

Magnitude production.—This method, suggested by S. S. Stevens (1958), but hitherto not extensively used to scale sensory magnitude, required *O* to exert forces whose apparent magnitudes were proportional to numbers supplied in succession by *E*—in these experiments, the numbers 3, 6, 10, 20, and 30.

The instructions were: "You are asked to begin by exerting a *moderate* pull on a handle. Let the number 10 stand for the apparent value of the force you exert. You will then be asked to exert forces in proportion to other numbers. For example, if given the number 25, you would attempt to exert an apparent force 2½ times as great as the force called 10."

In the first of two experiments each of 10 *O*s pulled the movable handle 50 times, 10 times for each number. The order in which the numbers were given was irregular and different for each *O*. In the second experiment, each of 10 *O*s, of whom 7 had taken part in the first experiment, pulled the stationary handle 35 times, 7 times for each number.

Magnitude estimation.—In two experiments, *O* was instructed (*a*) to increase the force applied to the handle until *E* signaled that a proper force had been achieved, (*b*) to let the number 10 stand for the apparent magnitude of that force, and (*c*) to assign numbers proportional to the apparent magnitude of each of a series of subsequent forces indicated by *E*'s signal. Since *O* occasionally "overshot" the force required, it was sometimes necessary for him to moderate the force by instructions from *E*.

Using the movable handle, nine *O*s estimated the apparent magnitude of six specified forces between 5 and 35 lb. The initial pull, called 10, equaled 17 lb. Sixty estimates were made in a single session, 10 at each of the six levels of force. Using the stationary handle, 10 *O*s, of whom 5 had already participated in the first experiment, estimated the magnitude of six levels of force, between 4 and 40 lb. The initial pull, called 10, equaled 15 lb., and *O* estimated each level six times. The *E* specified the levels in several irregular orders.

Category production.—These instructions were: "You are asked first to exert a light force on the handle, then a heavy force, and following these, forces of various magnitudes. The idea is to exert forces that belong to one of seven equal categories. The first force

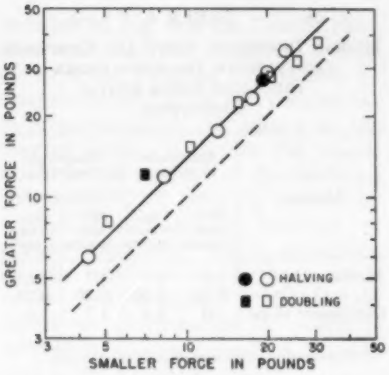


FIG. 2. Ratio production of forces exerted on movable handle. *Unfilled circles:* derived from median ratios of 150 one-half judgments (10 *Os* distributed among six equal intervals of force; see text). *Unfilled squares:* derived from median ratios of 150 double judgments (10 *Os*) distributed among six intervals of force. Dashed line shows equal physical force. Exponent of power function is 2.0. *Filled circle and square:* mean results of two *Os* (Fullerton & Cattell, 1892).

(the light one) will belong to Category 1; the second force (the heavy one) will belong to Category 7. If you are then given the number 5, for example, you will exert a force that belongs to the fifth category, *etc.*"

Results

Ratio production.—On the average, the ratio of the greater to the smaller force of a pair in 2 to 1 subjective relation turned out to be approximately constant at about 1.4/1.0. The largest deviation from this ratio occurred in the direction of smaller ratios when *O* attempted to double a rather large initial force.

Since for each judgment *O* determined the value of the first of a pair of pulls, the problem arises, how to average the data when different *Os* started from different initial values. Actually, several alternative procedures were found to give essentially the same result. One of these was to divide the entire range of initial forces into six equal intervals of force, and to obtain the median ratio (of the greater to the lesser force) belonging to each interval. Figure 2 shows the results in log-log coordinates. For each interval, the value of the greater force (ordinate) has been plotted against the value of the smaller force (abscissa). The fact that the data can be fitted reasonably well by a line with a slope of 1.0 signifies that the greater force tends to be a constant multiple of the smaller, and consequently that subjective magnitude approximates a power function of the force exerted. The position of the line suggests that apparent force grows roughly as the square of the force exerted.

Magnitude production (group results).—Figure 3 (top) records the group results obtained with the two kinds of apparatus. From the fact that the points (medians) determine straight lines in log-log coordinates, we may conclude that apparent force grows as a power of the force exerted. The similarity of the two functions in Fig. 3 suggests that there is a negligible difference between exerting force across a fairly short but perceptible distance and exerting force on a virtually stationary handle.

The lines of Fig. 3 have slopes close to 2.0; thus the results of magnitude production, like those of ratio production, imply that apparent force grows approximately as the square of the physical force.

Magnitude estimation (group results).—Figure 3 (lower) shows two functions determined by magnitude estimation. The linearity of these functions substantiates the outcome of ratio production and magnitude production, i.e., that the apparent magnitude grows as a power of exerted force, but the estimation functions are governed by a slightly smaller

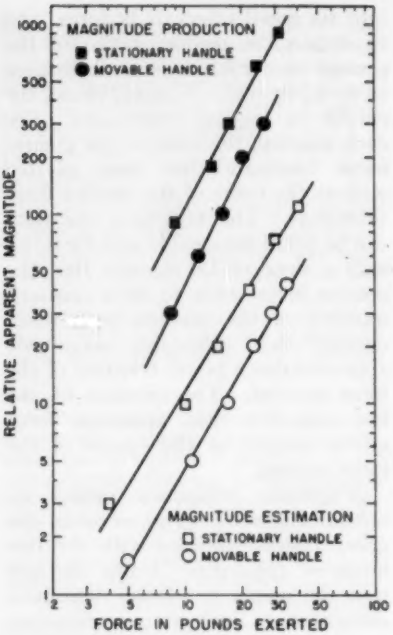


FIG. 3. Magnitude production and magnitude estimation (group results). The points stand for the median force exerted (filled points) or for the median estimate given (unfilled points). In order to separate the curves, the ordinate values of the respective functions (reading from top to bottom) have been multiplied by 30, 10, 2, and 1.

exponent—about 1.6 or 1.7, as inferred from the slopes of the lines in Fig. 3.

Individual results.—It turned out that the medians of the results for each *O* separately could be fitted well by straight lines in log-log coordinates. Although the slope varied from one *O* to another, the judgments of each individual *O* determined a good approximation to a power function. This was true for the results of both magnitude estimation and magnitude production. The median slope of all the individual functions is 1.7, but the tendency shown in Fig. 3 for magnitude production to yield steeper slopes than magnitude estimation also characterized the individual functions. As shown in Table 1, the median of the individual slope values, obtained under a particular experimental condition, agrees well with the slope value determined by the composite group results for the same condition. Also listed in Table 1 are the semi-interquartile ranges of the individual slope values.

TABLE 1

MEDIAN INDIVIDUAL SLOPE AND COMPOSITE GROUP SLOPE OBTAINED UNDER DIFFERENT EXPERIMENTAL CONDITIONS

Measure	Magnitude Production		Magnitude Estimation	
	Movable Handle	Stationary Handle	Movable Handle	Stationary Handle
Median slope	2.0	1.7	1.7	1.65
$(Q_3 - Q_1)/2$[a]	0.45	0.40	0.20	0.25
Composite slope	2.0	1.9	1.7	1.6

[a] Semi-interquartile range of individual slopes.

Category production.—The category scale of the forces exerted by the eight *O*s turned out to be nonlinear when plotted against the ratio scale determined by magnitude production. The

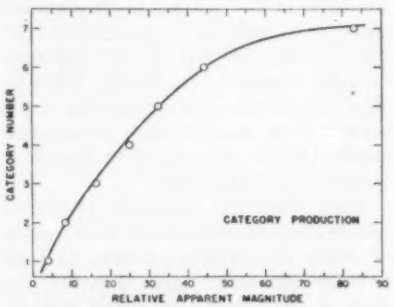


FIG. 4. Category production. Each point is the median of 56 judgments by eight *O*s. The ordinate is the category number specified by *E;* the abscissa is proportional to the square of the force exerted by *O*.

ordinate of Fig. 4 is the 7-point category scale, and the abscissa is a scale of apparent force as determined by magnitude production, i.e., the abscissa is proportional to the square of the exerted force. Although the points stand for the medians of the combined judgments of all eight *O*s, each *O* by himself determined a function of similar curvature. For a discussion of the nonlinearity between ratio scales and category scales, which seems to characterize all prothetic perceptual continua, see S. S. Stevens and E. H. Galanter (1957).

Discussion

Although the results show that the apparent magnitude of handgrip grows as a power function of the force exerted, in general, the production procedures yield slightly steeper slopes (larger exponents) than the estimation procedure. It has been suggested by S. S. Stevens (1958) that biases associated with the two procedures probably operate in opposite directions. Hence it may prove advisable to take seriously the results of both procedures and to assume that together they yield a more valid function than they do separately. Since the median exponent of all the individual functions equals 1.7, we shall probably not go far astray in concluding that the apparent magnitude grows approximately as the 1.7 power of the force exerted.

The variability among the slopes of the individual functions may conceivably come about in part because of anatomical differences among *O*s, e.g., in the capacity of the musculature or in the shape of the hands, but it appears doubtful that such differences constitute a very large source of the variability. If they did, we would expect an *O* who produced a steep function under one condition to produce a steep function under another condition as well. Actually, it turned out that *O*s who took part in more than one experiment did not produce highly consistent slopes. It would seem probable, therefore, that the variation in slope comes about primarily because, like any other performance, *O*'s attempt to fix upon and hold to a constant "quantitative" criterion varies from time to time, as well as from person to person.

Whatever the reason for the individual differences in slope may be, it is encouraging that the results obtained from the typical *O* agree fairly well with the early fractionation results reported by Fullerton and Cattell (1892). In that experiment, each of two *O*s made 100 attempts to double a force of about 3 kgm. and to halve a force of about 12 kgm. The mean forces produced by the two *O*s appear in Fig. 2.

It is also of interest that some results by Girden[2] seem to corroborate the present study. Three *O*s applied rotational pressure on a knob, first to a standard magnitude and then to a magnitude that appeared one-half as great. On the average, the ratio of the greater to the smaller force equaled approximately 1.5/1.0 (compared with 1.4/1.0 for the "halving" judgments in the present study). Although Girden's standard forces, which ranged from 50 to 200 gm., were orders of magnitude smaller than the forces judged in the present experiments, and although the nature of the muscular responses was considerably different, the results produced a power function with an exponent of about 1.7.

Some results reported by Bernyer (1959) are also of interest. The *O*s pushed on a lever, first to some standard magnitude (between 100 and 1000 gm.) and then to a magnitude that appeared one-half or twice as great. For each *O* the ratio of the greater to the lesser push was approximately constant at 1.4/1.0, the same ratio that was, on the average, obtained in the present study using force of handgrip.

Force Applied to the Skin

The perception of force applied to the passive skin is rather different

[2] Girden, E. Personal communication, 1958.

from the perception of force exerted actively by the muscles. The exertion of force excites the kinesthetic sense as well as the senses of touch and pressure, whereas forces received in a passive way presumably involve the senses of touch and pressure. It is not surprising, therefore, that the scales of subjective magnitude are different for these two kinds of stimulation. The following experiments show that the apparent magnitude of force applied to the palm of the hand grows approximately as the first power of the physical magnitude (as contrasted with the 1.7 power for muscular pull).

Procedure

The *O* rested his right hand and arm on a table, palm up. Forces on the palm were applied by means of the push rod of a Pelouze Tension Testing Scale. The *O* kept his eyes closed while *E* applied the desired level of force to *O*'s palm for about 2.5 sec. The stimuli were five levels of force, between 0.5 and 5 lb. It was possible to adjust the areal contact between the push rod and the skin by attaching smooth, circular brass contact buttons to the end of the rod. The diameters of the buttons used in three different experiments were 0.19, 0.50, and 1.25 in.

Except for the differences in the area over which the force was applied, the three experiments followed the same procedure. The instructions were to let the number 10 stand for the apparent magnitude of the first force (always 2 lb.), and to assign numbers to successive stimuli in proportion to apparent magnitude. After the presentation of the first stimulus, each *O* received the set of stimuli four times. Between the second and third sets, *E* presented again the 2-lb. stimulus, and identified it again as 10. The stimuli were presented in a different order to each *O*.

Each experiment employed 12 *O*s. Some took part in all three experiments, others in two of the three experiments, and still others in only one experiment.

Results

As shown in Fig. 5, the median estimates determine linear functions in log-log coordinates. The slopes

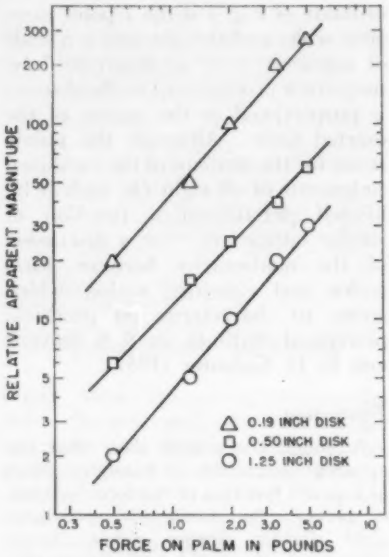


FIG. 5. Magnitude estimation of force applied to the palm of the hand. Each point stands for the median of 48 judgments by 12 *O*s. In order to separate the three functions the medians of the top function have been multiplied by 10 and the medians of the middle function have been multiplied by 2.

of these functions are all slightly larger than unity—about 1.1. The area of contact did not seem to make much difference. This outcome indicates that the apparent magnitude of force applied to the skin grows as a power function of the stimulus, with an exponent that is decidedly smaller than the exponent that governs the scale for muscular pull.

CROSS-MODAL MATCHING BETWEEN FORCE OF HANDGRIP AND FORCE APPLIED TO THE SKIN

A procedure has recently been described by S. S. Stevens (1959a) which makes it possible to verify the relative forms of ratio scales of

apparent magnitude. Under this procedure, *O* tries to match the apparent magnitudes of sensations aroused through different sense modalities. If two scales are both power functions of the stimulus, direct cross-modal matching ought to determine an equal-sensation function that plots as a straight line in log-log coordinates. The slope of the line should equal the ratio between the exponents of the separate functions. By cross-modal matching, S. S. Stevens (1959a) was able to validate the subjective scales for loudness, vibration, and electric shock, as well as the scales for vibration applied to two different parts of the body (1959b).

A similar matching procedure was used to compare the scale for force applied to the passive palm with the scale for active muscular pull. If, for the actively exerted force, the apparent magnitude f_a relates to the physical force F_a by the power function:

$$f_a = (F_a)^{1.7} \qquad [1]$$

and if the apparent magnitude f_p of force applied to the palm relates to the physical force F_p by the power function:

$$f_p = (F_p)^{1.1} \qquad [2]$$

then it follows that the equal-magnitude function, determined by direct matches, ought to have the form:

$$F_a = (F_p)^{0.65}. \qquad [3]$$

The exact value of the exponent to be expected will depend, of course, upon procedural details, but an exponent in the vicinity of 0.6 or 0.7 is a reasonable expectation.

Procedure

The *O* exerted forces on the dynamometer (stationary handle) to match the apparent values of forces applied to his left palm by *E*.

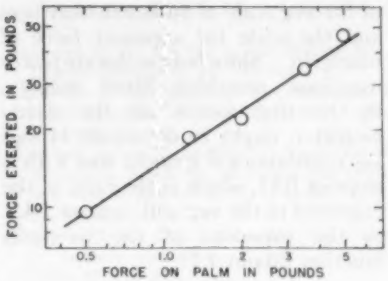


FIG. 6. Equal-sensation function relating active and passive force. The force exerted on the dynamometer by the right hand is plotted against the force applied to the palm of the left hand. Each point stands for the median of 40 matches by 10 *O*s.

The *O* made four matches of each of five levels of force between 0.5 and 5 lb. Ten *O*s took part. Before the session, *O* was told to approach a final judgment by bracketing.

Results

Most *O*s reported that the judgment was not particularly difficult. As might be expected, the *absolute* value of the force exerted for a match varied considerably from one *O* to another. Our particular interest, however, is in the *relative* values, for these determine the *form* of the equal-sensation function. In Fig. 6, the median of the force exerted by the 10 *O*s is plotted against the force applied to the palm. In log-log coordinates the medians determine a straight line whose slope, 0.66, is close to the slope predicted. Thus, to a fair approximation, the equal-sensation function confirms the difference in size between the exponents obtained in the direct ratio-scaling experiments.

CROSS-MODAL MATCHING BETWEEN FORCE OF HANDGRIP AND LIFTED WEIGHTS

The purpose here was to extend the matching procedure to a comparison

of the veg scale of apparent heaviness and the scale for apparent force of handgrip. Since both scales are power functions, matching lifted weights, by exerting forces on the dynamometer, ought to determine in log-log coordinates a straight line with a slope of 0.85, which is the ratio of the exponent of the veg scale (about 1.45) to the exponent of the handgrip function (about 1.7).

Procedure

Using his right hand, *O* exerted forces on the stationary handle to match the apparent heaviness of weights lifted simultaneously with the left hand. Each of six weights, ranging from 28 to 480 gm., was matched three times by each of 10 *O*s.

Results

Figure 7 shows the median forces exerted on the hand dynamometer to match the lifted weights. The straight line in log-log coordinates has a slope of 0.8, which is reasonably close to the predicted value of 0.85. This outcome suggests that the veg scale and the handgrip scale are, in fact, slightly different functions of physical force. As a further check on the reliability of this difference, seven of the *O*s who participated in this experiment later made magnitude estimates of the same set of weights. The slope of the power function determined by these estimates turned out to be 1.4, which is reasonably consistent both with the veg scale and with the value that would be predicted from a knowledge of equal-sensation function and the exponent for apparent force of handgrip.

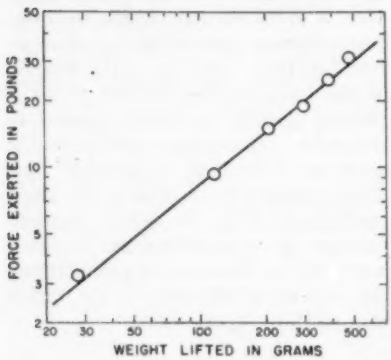


FIG. 7. Equal-sensation function between handgrip and lifted weights. The force exerted by the right hand is plotted against the value of the weight lifted by the left hand. Each point stands for the median of 30 matches by 10 *O*s.

DISCUSSION

The foregoing series of experiments has shown not only that a consistent scale of apparent force can be constructed by different scaling procedures, but that the validity of the resulting scale can be checked by cross-modal comparisons with other scales, independently arrived at. The outcome of these comparisons adds further evidence of the generality of the power function, and gives added assurance that we know the approximate values of the exponents of the various subjective scales.

SUMMARY

By three types of direct scaling procedures, it was demonstrated that the apparent magnitude of handgrip grows approximately as a power function of the force exerted on a hand dynamometer. The exponent of the function is about 1.7. The exponent of the power functions for some individual *O*s may depart considerably from the median value.

It was also shown that the apparent magnitude of a force applied to the passive skin grows roughly proportionally to the physical force. The results of magnitude estimation show that the apparent magnitude grows approximately as the 1.1 power of the force applied to the palm of the hand.

In order to test the validity of these two psychophysical functions, *O*s were asked to exert forces on the dynamometer that appeared to match the magnitude of forces applied to the palm. This procedure determined an equal-sensation function, whose

form is a power function, and whose slope is predicted by the ratio between the exponents of the two subjective scales.

The matching procedure was also used to compare the scale of apparent handgrip with the veg scale of apparent heaviness. The *O* produced forces on the dynamometer that appeared to match the heaviness of various lifted weights. The equal-sensation function thus generated verifies the fact that the exponent of the veg scale is about 80% as large as the exponent of the handgrip function.

References

BERNYER, G. Discussion in symposium: Le probleme de la mesure. *Proc. 15th Int. Congr. Psychol., Brussels*, 1957. Pp. 99–100.

FULLERTON, G. S., & CATTELL, J. McK. *On the perception of small differences.* Philadelphia: Univer. Pennsylvania Press, 1892.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

STEVENS, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.

STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, **57**, 201–209. (a)

STEVENS, S. S. Tactile vibration: Dynamics of sensory intensity. *J. exp. Psychol.*, 1959, **57**, 210–218. (b)

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received December 31, 1958)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

SUPPLEMENTARY REPORT: MAINTENANCE OF A PREVIOUSLY LEARNED CONCEPT AS A FUNCTION OF PARTIAL REINFORCEMENT

ARNOLD H. BUSS

University of Pittsburgh

A previous study (Buss, 1953) attributed the slowness of nonreversal shift to partial reinforcement of the previously acquired concept. In neither that study nor subsequent ones was the percentage of partial reinforcement varied systematically. Gormezano and Grant (1958) varied the percentage of reinforcement of partially correlated concepts during the *learning of the original concepts*. The present experiment varies partial reinforcement of the previous concept after a shift is required, during *learning of the second concept*. If the original analysis is correct, the greater the percentage of reinforcement, the more the previously learned concept will be maintained.

Method.—Eighty college students from introductory psychology classes were randomly assigned to four groups of 20 each. The stimuli were the same wooden blocks used in the previous study. There were two heights, three shapes, three colors, and three top surface areas, and each block differed from every other one in at least one of these four attributes.

There were two series of stimuli. Series 1 consisted of 10 stimuli, and all groups learned a shape discrimination (circles positive, noncircles negative) with 100% reinforcement. Series 2 consisted of 24 stimuli, and a height discrimination (tall positive, short negative) was 100% reinforced for all groups. In Series 2 the shape discrimination previously learned received 0%, 25%, 50%, and 75% reinforcement in Groups I–IV, respectively. The 25% reinforcement (Group II) of the shape discrimination was accomplished by manipulating height and shape combinations. Two of the 8 circles were tall, and 4 of the 16 noncircles were short. This partial covariation insured that *S* would be reinforced 25% of the time for shape while being reinforced 100% of the time for a height discrimination. The covariation between height and shape was 50% in Group III and 75% in Group IV. In Group I (0%), all tall stimuli were noncircles and all short stimuli were circles. Thus there were 0% reinforcement of circles positive, noncircles negative and 100% reinforcement of a reverse shape discrimination (noncircles positive, circles negative).

In both series the stimuli were presented singly. The *S* said "Vec" or "not Vec," after which the *E* said *Right* or *Wrong*.

Results and Discussion.—The measure was the frequency of "Vec" responses. In Series 1, *S*s in all groups learned the shape concept, making "Vec" responses to circles and "not Vec" responses to noncircles (squares and triangles). In Series 2 this shape concept was reinforced with varying percentages, depending upon the group. Series 2 was divided into halves, and the percentage frequencies of "Vec" responses are shown in Table 1.

During the first half of Series 2, Groups I and II were in the process of shifting from the previously learned shape concept to the height concept that was now being 100% reinforced. Groups III and IV were maintaining the previously learned shape concept.

In the second half of Series 2, Group I shifted to the reversed shape concept that was being reinforced simultaneously (and 100%) with the height concept. Group II now failed to discriminate between the shapes of the stimuli, and Groups III and IV continued to maintain the shape concept that had been learned in Series 1.

"Vec" responses to circles show a steady increase and to noncircles a steady decrease as a function of percentage of reinforcement of the previously learned shape concept. Analyses of variance of "Vec" response frequencies yielded *F*s of 22.4 ($df = 3, 76$;

TABLE 1

PERCENTAGE OF "VEC" RESPONSES IN SERIES 2

Group	First Half		Second Half	
	Circle	Non-circle	Circle	Non-circle
I (0%)	38	43	28	82
II (25%)	45	43	50	63
III (50%)	65	26	76	40
IV (75%)	86	19	95	10

$P = .001$) and 37.9 ($df = 3, 76$; $P = .001$) for circles and noncircles, respectively.

Thus, the greater the percentage of reinforcement, the more the previously acquired concept is maintained. This finding confirms the original formulation (Buss, 1953) and is consistent with the Gormezano and Grant (1958) results with the Wisconsin Card Sorting Test. They demonstrated that partial reinforcement of a concept during original learning impedes learning of a new response. The present study shows that partial reinforcement of the old concept during the learning of a new concept impedes its learning.

References

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

Gormezano, I., & Grant, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

(Received May 15, 1959)

Journal of Experimental Psychology
Vol. 58, No. 5, 1959

REPLICATION REPORT: WORK AND REST AS VARIABLES IN CYCLICAL MOTOR PERFORMANCE

ABRAM M. BARCH

Michigan State University

Kimble and Bilodeau (1949) analyzed the separate and joint effects of work and rest variables using the turning task of the Minnesota Rate of Manipulation Test (MRMT). They found that lengthening the trial period produced an immediate and constant performance decrement. However, their 10-sec. and 30-sec. trial groups had a qualitative as well as quantitative difference in task activities.

Every trial for each group began with the upper right-hand corner of the board. More of the 30-sec. trial *Ss* could complete the top row of blocks and would need to reverse direction of movement and hand activity than the 10-sec. trial group *Ss*. This reversal of movement and hand activity is one of the sources of difficulty on the turning task, especially during the early trials. This methodological flaw could have produced the finding of an immediate and constant effect of increasing trial duration and might also be related to the lack of interaction found between work and rest variables.

Kimble and Bilodeau also reported that the increasing trial duration had a greater effect on performance than increasing rest duration. However, this difference was decreasing with practice and the relationship might have reversed with practice beyond 300 sec. This is of some importance in that Kimble (1949) applied his general analysis to the alphabet printing task and replicated all of the findings of the previous study, with the exception of the one concerned with the relative importance of work and rest. However, an error apparently was made in the analysis of this latter study. If the reader will analyze the data cited, he will find that the results show, with a total of 600 sec. of practice, that the effect of increasing rest duration increases with practice until it becomes equivalent to the effect of increasing trial duration.

Method.—The present study replicated the significant features of the Kimble and Bilodeau experiment with three exceptions: (*a*) *Ss* began each trial after the first at the point reached on the board at the end of the previous trial; (*b*) total amount of practice was increased from 300 to 600 sec.; (*c*) *Ss* were naive to the MRMT. Each group contained 18 men and 6 women from an introductory psychology class.

Results.—Performance increments over trials (Fig. 1) were still evident even after 9 min. of practice, while the Kimble and Bilodeau groups had apparently reached asymptotic performance within 5 min. In addition, our groups show greater irregularity in performance from one 30-sec. block to another. These differences are probably due to our use of naive *Ss* and the shifting starting point.

An analysis of variance (Lindquist, 1953, Type III) was used to test the interaction of work and rest over 30-sec. trials. The effect of increasing trial duration was highly significant ($F = 7.52$ for 1 and 92 df) and did not interact with amount of practice on the task or with the effect of increasing rest duration. The effect of increasing rest increased with practice. Thus, all the statistical findings of the previous study were supported despite procedural changes.

Figure 2 presents the graphical analysis of the relative effects of changing work

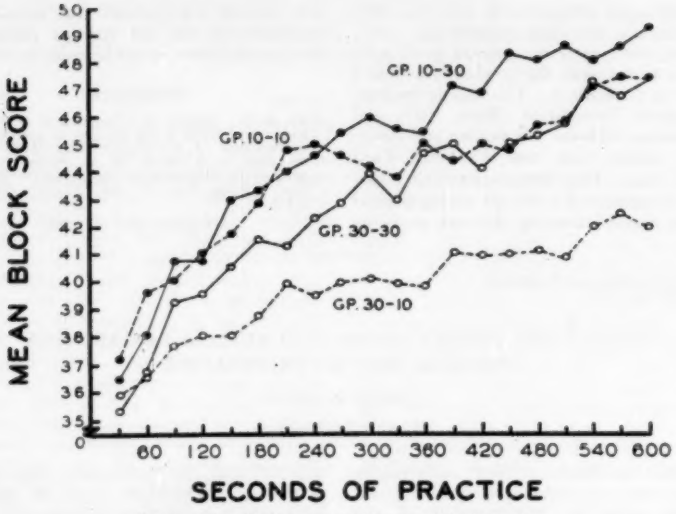


FIG. 1. Mean block score over 30-sec. periods of practice.

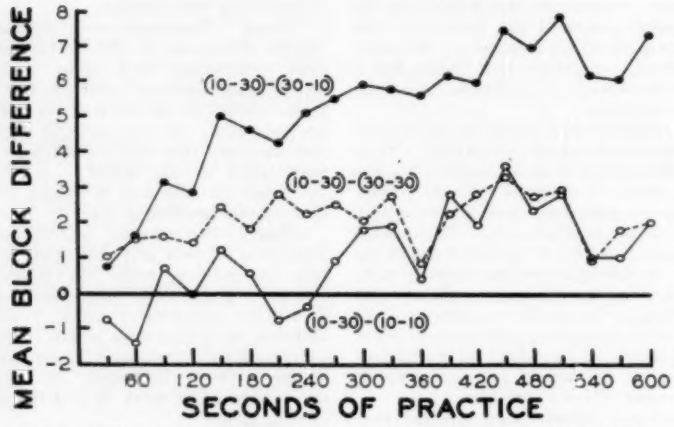


FIG. 2. Graphical analysis of single and joint effects of increasing work and rest duration.

period, rest period, or both. Early in practice the score difference attributable to lengthening trial duration is definitely greater than that due to shortening the rest period. As practice continues these two differences become equivalent but do not reverse their position. This result is the same as that obtained for the alphabet printing task when the cited data is used (Kimble, 1949) and probably would have been obtained in the original MRMT study if more practice had been given.

The marked agreement found in the three studies on work and rest variables would seem to demand theoretical attention.

REFERENCES

KIMBLE, G. A. A further analysis of the variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 332–337.

KIMBLE, G. A., & BILODEAU, E. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.

LINDQUIST, E. G. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received June 11, 1959)

LOOKING FOR A BOOK PUBLISHER?

It is no secret that publication is regarded as the foundation stone of a scholar's career. In order to gain recognition, prestige, and advancement, the scholar must seek publication.

Exposition Press offers scholars a complete publishing service, under our special academic imprint, *Exposition-University Books*. Though the proportion of scholarly publications in the lists of the trade houses is shrinking, due to economic pressures, several of our recent titles in the academic fields have gained popular reviews, professional recognition and steady sales. These books are the result of a publishing plan that is bringing a steady flow of scholarly writers to Exposition Press. They reflect the high editorial standards and quality of design and production which have won the respect of libraries, schools, booksellers and critics.

FREE! The behind-the-scenes story of book publishing revealed in two fact-filled, illustrated brochures, containing a detailed description of our 40%-royalty subsidy plan including a breakdown of contract terms and typical costs (in print for the first time). Copies are available on request. Your inquiries and manuscripts are invited. An editorial appraisal will be furnished promptly without obligation.

Please write to Jarold Paine, Editorial Department.

EXPOSITION PRESS, *Main Office, 386 Park Ave. S., N.Y. 16, N.Y.*

PSYCHOLOGICAL MONOGRAPHS

The monographs listed below are still in print—but for how long?*

ADELBERT AMES, JR. *Visual Perception and the Rotating Trapezoidal Window.* No. 324. *$1.00*

JANE SHOHL. *Effects of Exposure to Sound on Discrimination Performance in the Rat.* No. 329. *$1.00*

PHILIP WORCHEL. *Space Perception and Orientation in the Blind.* No. 332. *$1.00*

P. R. WENDT. *Development of an Eye Camera for Use with Motion Pictures.* No. 339. *$1.00*

MONCRIEFF SMITH and EDNA A. WILSON. *A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer.* No. 359. *$1.50*

M. R. ROSENZWEIG and W. A. ROSENBLITH. *Responses to Auditory Stimuli at the Cochlea and at the Auditory Cortex.* No. 363. *$1.00*

M. B. JONES. *An Experimental Study of Extinction.* No. 369. *$1.00*

G. F. RABIDEAU. *Differences in Visual Acuity. Measurements Obtained with Different Types of Targets.* No. 395. *$1.00*

---

* A selection from Volumes 65–69. (Write for complete list of in-print items.)

Order *now* from: **American Psychological Association**
1333 Sixteenth Street, N. W.
Washington 6, D. C.